

UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE 564 :
PHYSIQUE EN ÎLE-DE-FRANCE

LABORATOIRE DE PHYSIQUE THÉORIQUE ET MODÈLES STATISTIQUES

DISCIPLINE : PHYSIQUE

THÈSE DE DOCTORAT

Soutenue le 22 juin 2015 par

Anthony Perret

**Statistique d'extrêmes
de variables aléatoires fortement corrélées**

Directeur de thèse : M. Grégory Schehr

Chargé de recherche (Paris-Sud)

Composition du jury :

Rapporteurs : M. Gernot Akemann

Professeur (Bielefeld)

M. Malte Henkel

Professeur (Nancy)

Examineurs :

M. Tom Claeys

Chargé de cours (Louvain la Neuve)

M. Henk Hilhorst

Professeur (Paris-Sud)

M. Henri Orland

Chercheur CEA (Saclay)

M. Clément Sire

Directeur de recherche (Toulouse)

Table des matières

Introduction	9
Extrêmes de marches aléatoires	10
Valeurs propres maximales de matrices aléatoires	11
Densité à partir du maximum	12
Organisation et résultats principaux	14
Notations et formules	16
1 Statistique d'extrêmes de variables aléatoires indépendantes	19
1.1 Statistique du maximum	19
1.1.1 la classe de Gumbel	20
1.1.2 la classe de Fréchet	21
1.1.3 la classe de Weibull	21
1.1.4 Remarques	22
1.2 Densité moyenne à partir du maximum	22
1.3 Premier gap et statistiques d'ordre	24
I Mouvement brownien	27
2 Introduction	29
2.1 Le mouvement brownien	29
2.1.1 Présentation	29
2.1.2 Définition	29
2.2 Les browniens contraints	30
2.3 Statistique d'extrême du mouvement brownien	31
2.4 La densité à partir du maximum	33
3 La méthode des propagateurs	35
3.1 Propagateurs	35
3.2 Illustration de la méthode : calcul de $\langle \rho_{\max}(r, t) \rangle$ dans le cas du brownien libre	37
3.3 Calcul des différents moments $\langle \rho_{\max, B}^k(r, t) \rangle$ pour le pont et l'excursion . . .	39
3.4 Résultats pour les autres browniens contraints	41
4 Extensions : étude temporelle et cas de plusieurs browniens	45
4.1 Étude temporelle	45
4.1.1 Cas du brownien libre	46
4.1.2 Cas du pont	47
4.1.3 Comportements et résultats numériques	48
4.2 Cas de plusieurs browniens	49
4.2.1 Plusieurs browniens indépendants	49

4.2.2	Ponts répulsifs	53
5	La méthode de l'intégrale de chemin	59
5.1	Principe	59
5.1.1	Brownien libre	59
5.1.2	Pont	62
5.2	Applications à l'étude de certaines fonctionnelles du maximum	63
5.2.1	Statistique complète de la densité à partir du maximum	63
5.2.2	Applications aux cas où $V(x) = x^\alpha$	70
5.2.3	Cas exactement soluble du potentiel $V(x) = 1/x$	79
5.2.4	Cas du brownien libre	79
	Appendices	83
A	La famille de fonctions $\Phi^{(j)}$	85
B	Divers calculs dans le cas de plusieurs ponts indépendants	87
B.1	Calcul de la densité à partir du maximum pour n ponts indépendants . . .	87
B.2	Calcul de $\langle r \rangle$ dans la limite d'un grand nombre de ponts indépendants . . .	88
B.3	Limite thermodynamique d'un grand nombre de ponts indépendants	88
C	Cas de plusieurs ponts répulsifs	91
C.1	Génération de ponts répulsifs : mouvement brownien de Dyson	91
C.2	Temps local des ponts répulsifs dans la limite thermodynamique de grand n	91
D	Moments de la densité à partir du maximum du pont	93
E	Recherche en environnement inconnu : Algorithme d'Odlyzko	95
II	Matrices Aléatoires	99
6	Introduction	101
6.1	Les matrices aléatoires	101
6.2	Les ensembles gaussiens	101
6.3	Densité des valeurs propres	102
6.4	Loi de Tracy-Widom	105
6.5	Transition de phase associée à λ_{\max} : le modèle de May	108
7	Statistique à partir de la valeur propre maximale pour l'ensemble gaussien unitaire	111
7.1	Densité à partir de la valeur propre maximale et gap entre les deux plus grandes valeurs propres	111
7.2	Formule à N fini	113
7.3	Limite à grand N	116
7.3.1	Régime $r \propto \sqrt{N}$: régime de volume	116
7.3.2	Régime $r = \mathcal{O}(N^{-1/6})$: régime de bord	117
7.4	Simulations numériques	121
7.4.1	Simulation rapide des matrices gaussiennes	121
7.4.2	Évaluation numérique des fonctions d'échelle	122

8	Généralisations et Applications	125
8.1	Ouvertures aux ensembles gaussiens β	125
8.2	Application à la minimisation d'une forme quadratique	128
8.3	Applications à un modèle de verre de spin en champ moyen	129
8.4	Densité à partir du maximum et premier gap à λ_{\max} fixée	130
9	Statistique d'extrêmes pour les matrices de Wishart	133
9.1	Modèle des matrices de Wishart	133
9.2	Polynômes orthogonaux	135
9.3	Équation de Painlevé V à N fini	137
9.4	Étude du <i>hard edge</i>	138
9.5	Étude du <i>soft edge</i>	141
	Appendices	143
F	Analyse du régime de bord pour des matrices gaussiennes unitaires	145
G	Densité à partir du maximum et premier gap pour des matrices de Wishart $a = 0$	149
G.1	Modèle de Wishart pour $a = 0$	149
G.2	Densité de valeurs propres	150
G.3	Densité à partir de la valeur propre minimale	151
G.4	Statistique du gap entre les deux plus petites valeurs propres	152
	Conclusion et perspectives	155
	Remerciements	159
	Bibliographie	170
	Publications	171
	<i>Near-extreme statistics of brownian motion.</i>	171
	<i>Near-extreme eigenvalues and the first gap of hermitian random matrices.</i> . . .	177
	<i>The density of eigenvalues seen from the soft edge of random matrices in the Gaussian β-ensembles.</i>	207
	<i>On certain functionals of the maximum of Brownian motion and their applications.</i>	224
	<i>Finite N corrections to the limiting distribution of the smallest eigenvalue of Wishart complex matrices.</i>	263

Introduction

Dans de nombreuses situations, les événements extrêmes, peuvent avoir des conséquences dramatiques et il est donc nécessaire de les étudier précisément. On peut donner une multitude d'exemples comme l'étude des grandes catastrophes naturelles : crues [43, 95], tempêtes [9], incendies [3], tremblements de terre [55, 120, 160]..., des records [48, 49], mais aussi en finance avec les crises financières dues aux fluctuations du marché boursier [11, 15, 53, 104, 129] ou encore en résistance des matériaux [7, 33]... C'est dans ce contexte qu'intervient l'étude des valeurs extrêmes [77] c'est-à-dire l'étude des événements proches des événements maximum ou minimum. Si on se donne une collection de N variables aléatoires $\{x_i | i = 1, \dots, N\}$ de loi quelconque, il est alors intéressant d'étudier la densité de probabilité du maximum $x_{\max} = \max\{x_i | i = 1, \dots, N\}$

$$p_{\max}(x, N) = \text{Pr.}[x_{\max} < x], \quad (1)$$

et en particulier son comportement lorsque l'on s'intéresse à un grand nombre de variables aléatoires c'est-à-dire dans la limite thermodynamique de grand N . De façon analogue, il est intéressant d'étudier la distribution de la variable minimale $x_{\min} = \min\{x_i | i = 1, \dots, N\}$. Récemment, plusieurs travaux ont montré que les statistiques d'extrêmes jouent également un rôle important en physique statistique [14, 40, 99]. En effet, si on suppose la collection $\{-x_i | i = 1, \dots, N\}$ comme étant celle des énergies possibles - des états - pour un système physique, l'état du système à très basses températures est gouverné par l'état fondamental c'est-à-dire l'état d'énergie minimale $E_{\min} = -x_{\max}$ [58, 59, 116]. L'étude des extrêmes est donc également importante dans les problèmes d'optimisation où l'on recherche cet état de plus basse énergie. Par ailleurs, la dynamique de tels systèmes est gouvernée, aux temps longs et à basse température, par les plus grandes barrières de potentiel, le temps de relaxation de ces systèmes physiques tels que les verres de spin est donc directement relié à l'étude des valeurs extrêmes [14]. Le cas des variables aléatoires indépendantes et identiquement distribuées est désormais bien compris grâce aux travaux [61, 71, 76, 77, 78, 139] dont je discuterai les principaux résultats dans le chapitre suivant. Il a été prouvé que la distribution du maximum p_{\max} dans la limite de grand N , après renormalisation et centrage, ne peut tendre que vers trois différentes lois, trois classes d'universalités, que sont les lois de Gumbel, de Fréchet et de Weibull. Dans la plupart des systèmes, on est néanmoins contraint d'étudier des collections de variables aléatoires corrélées. Dans le cas où celles-ci sont faiblement corrélées - la longueur de corrélation ξ étant très petite devant le nombre N de variables aléatoires - on peut se ramener aux cas des variables aléatoires indépendantes. En effet, en décomposant les N variables en paquets de ξ variables (qui sont donc essentiellement indépendants entre eux), le maximum global peut alors être vu comme le maximum des maxima locaux de chaque paquet qui sont alors des variables indépendantes. Le cas difficile à traiter est donc le cas des variables fortement corrélées - c'est-à-dire quand la longueur de corrélation est de l'ordre du nombre de variables aléatoires - mais c'est aussi le cas qui arrive dans la plupart des systèmes physiques comme ceux cités au début de cette introduction. Un exemple de modèle pour

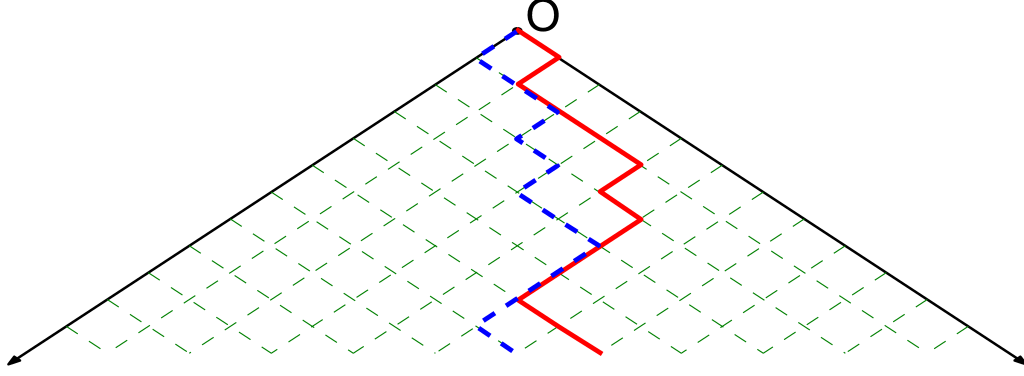


FIGURE 1: Deux exemples de chemins de 12 pas pour un polymère dirigé en milieu aléatoire. Le coût énergétique de chaque chemin est la somme des énergies associées aux sites visités. À cause du recouvrement possible entre les chemins, les différentes énergies possibles sont fortement corrélées.

lequel il est nécessaire d'étudier des variables aléatoires fortement corrélées est celui du polymère dirigé en milieu aléatoire [86]. Ce modèle est un archétype simple de système désordonné qui a plusieurs applications comme le piégeage d'interface élastique [83], des modèles de croissance d'interface dans la classe d'universalité de Kardar-Parisi-Zhang [94], la turbulence de Burgers [69], les verres de spins [114] et des problèmes d'alignement de séquences d'ADN [23, 84]. Le polymère dirigé est une marche aléatoire de N pas sur un réseau carré \mathbb{N}^2 avec des sauts $(+1, 0)$ et $(0, +1)$ comme le montre la figure 1. On associe à chaque site $(i, j) \in \mathbb{N}^2$ un coût énergétique $\epsilon_{i,j}$ distribué indépendamment de site à site. Cette variable gelée introduit ainsi du désordre. L'énergie d'un chemin est donnée par la somme des énergies associées aux sites visités et l'on s'intéresse au chemin de plus basse énergie. À cause du recouvrement possible entre les différents chemins, comme le montre la figure 1, les niveaux d'énergie des différents chemins sont fortement corrélés. Ainsi, les fluctuations du chemin de plus basse énergie ne sont pas décrites par le cas des variables indépendantes [6, 19, 20, 90, 130, 131]. Pour le cas des variables fortement corrélées, il n'existe pas de théorie générale. Dans le but de mieux comprendre ce cas, il est nécessaire d'étudier en premier lieu des exemples à la fois simples, pour pouvoir trouver des résultats analytiques, mais aussi suffisamment riches. Deux pistes de modèles physiques de variables fortement corrélées pouvant être étudiés analytiquement se démarquent : les positions d'une **marche aléatoire** et les valeurs propres de **matrice aléatoire**. Ce sont les deux modèles que nous allons étudier dans cette thèse. Dans les deux paragraphes qui suivent, nous introduisons brièvement ces deux modèles.

Extrêmes de marches aléatoires

Une première façon de créer un ensemble de variables fortement corrélées est de considérer une marche aléatoire de N pas

$$\begin{cases} x_0 = 0 \\ x_i = x_{i-1} + \eta_i, \quad i \in [1, N], \end{cases} \quad (2)$$

où les longueurs de pas η_i sont identiquement distribuées suivant une loi de second moment σ^2 que l'on supposera fini, par exemple une loi gaussienne. Cet ensemble de variable est alors fortement corrélé $\langle x_i x_j \rangle = \min(i, j) \sigma^2$. Un exemple de marche de $N = 20$

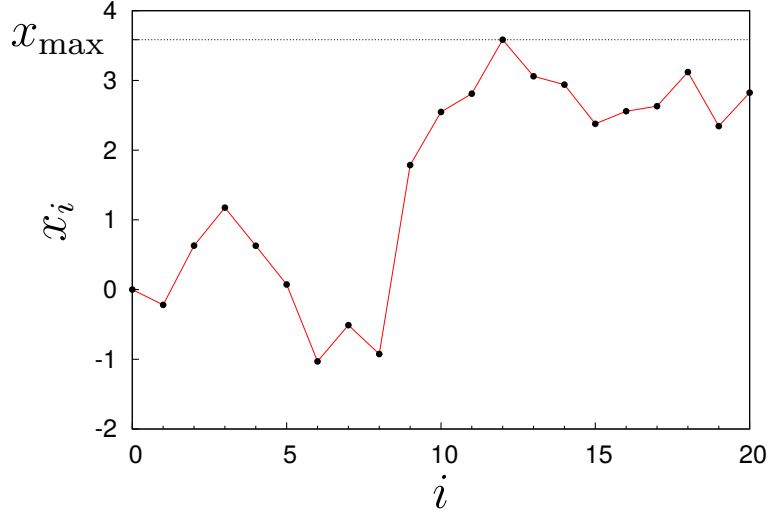


FIGURE 2: Exemple d'une marche aléatoire de 20 pas. $x_{\max} = \max\{x_i | i \in [0, N]\}$ est la position maximale du marcheur.

pas est tracée sur la figure 2 où les pas sont distribués suivant une loi normale centrée réduite. Il est possible d'étudier la statistique de la position maximale du marcheur $x_{\max} = \max\{x_i | i \in [0, N]\}$. Dans la limite thermodynamique de grand N qui nous intéresse, la marche aléatoire tend vers un mouvement brownien dont il est possible d'étudier analytiquement la statistique complète du maximum et même la loi jointe du maximum et de son temps d'atteinte [138]. De plus, comme nous le verrons par la suite, le mouvement brownien peut être facilement contraint, par exemple à rester positif ou encore d'être périodique [105], ce qui permet dans ce cas de modéliser plus précisément des collections de variables aléatoires corrélées. Le mouvement brownien et plus généralement les marches aléatoires possèdent de multiples applications comme par exemple la modélisation d'une interface élastique [103]. Dans ce contexte, la hauteur maximale relative de telle interface est régie par la loi d'Airy, que nous rediscuterons dans la section 5.2.2, et qui intervient dans de nombreux problèmes [34, 62, 106].

Valeurs propres maximales de matrices aléatoires

Une autre façon de créer un ensemble de variables fortement corrélées est de considérer les N valeurs propres d'une matrice aléatoire de taille $N \times N$ c'est-à-dire une matrice dont les éléments sont eux-mêmes des variables aléatoires (pouvant être éventuellement corrélées). Les matrices aléatoires, introduites en physique nucléaire par Wigner [163] pour modéliser les hamiltoniens très complexes des gros noyaux atomiques ont désormais trouvé de très nombreuses applications dans le domaine de la physique théorique en particulier à la physique nucléaire, au chaos quantique, aux systèmes désordonnés et à la théorie des nombres [113]. Diverses classes de matrices aléatoires ont été alors créées pour modéliser différentes collections de variables, en particulier suivant les symétries sous-jacentes des variables. Nous nous concentrerons par la suite uniquement sur les ensembles de matrices gaussiens et de Laguerre-Wishart qui font partie de la grande famille des ensembles invariants c'est-à-dire pour lesquels la loi de probabilité est invariante par action d'un groupe de symétrie. Pour les deux cas considérés, la distribution jointe des N valeurs propres est

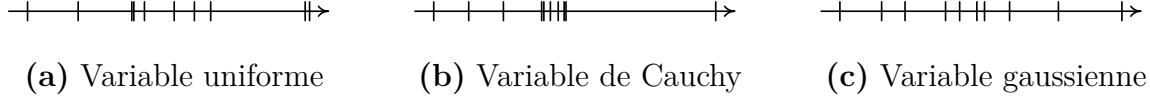


FIGURE 3: Répartitions typiques pour des variables aléatoires indépendantes et identiquement distribuées : **(a)** suivant une loi uniforme sur un segment, **(b)** suivant une loi de Cauchy, **(c)** suivant une loi gaussienne. Pour chacun des cas, on a généré 10 variables indépendantes que l'on a tracées avec une échelle pour avoir un même écart entre le minimum et le maximum obtenus dans chacun des trois cas. On remarque un comportement bien distinct aux alentours des extrêmes avec un regroupement dans le cas **(a)** de la loi uniforme alors que ceux-ci sont très éloignés dans le cas **(b)** d'une loi de Cauchy.

de la forme [163]

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_{\beta, N}} \prod_{i < j} (\lambda_i - \lambda_j)^\beta \exp\left(-\frac{\beta}{2} \sum_{i=1}^N V(\lambda_i)\right), \quad (3)$$

avec un potentiel $V(\lambda) = \lambda^2$ quadratique pour les matrices gaussiennes et un potentiel $V(\lambda) = \infty$ si $\lambda < 0$ et $V(\lambda) = \lambda$ sinon dans le cas des matrices de Wishart. Cette loi, que nous rediscuterons par la suite, montre bien que les N valeurs propres forment un ensemble de variables aléatoires fortement corrélées qui se repoussent entre elles à cause du déterminant de Vandermonde $\prod_{i < j} (\lambda_i - \lambda_j)$. Dans ce contexte, la loi de Tracy et Widom [154, 155] qui décrit les fluctuations typiques de la plus grande valeur propre dans la limite thermodynamique de grand N est un pilier de la théorie des statistiques d'extrêmes de variables aléatoires fortement corrélées. Nous rediscuterons plus en détails cette loi et ses différents applications dans la section 6.4.

Densité à partir du maximum

L'étude de la statistique du maximum, bien qu'intéressante, peut ne pas être suffisante. En reprenant un contexte de physique statistique où les variables $-x_i$ sont les énergies possibles d'un système physique, l'étude de la statistique du maximum permet de connaître l'état fondamental $E_{\min} = -x_{\max}$ gouvernant le système à basse température. Néanmoins, la thermodynamique à température $T > 0$ du système est principalement dominée par les états proches de cet état fondamental : cet état est-il très loin des autres ou, au contraire, est-il accompagné d'une multitude d'autres états "proches" pouvant être excités ? Cette variété de comportement est illustrée dans la figure 3 où l'on donne trois exemples de collection de 10 variables aléatoires : dans le cas **(b)**, les extrêmes sont très éloignés des autres variables aléatoires tandis que cela est moins prononcé dans le cas **(c)** et qu'il y a même un regroupement vers les valeurs extrêmes pour le cas **(a)**. Pour analyser ces différents comportements, nous introduisons la densité à partir des extrêmes [139] en imitant la densité. Pour une collection de N variables, il est naturel de définir la densité par

$$\rho_0(x, N) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i), \quad (4)$$

où δ est la distribution de Dirac. En reprenant le contexte de physique statistique où les variables $-x_i$ sont les énergies possibles d'un système physique, la densité moyenne

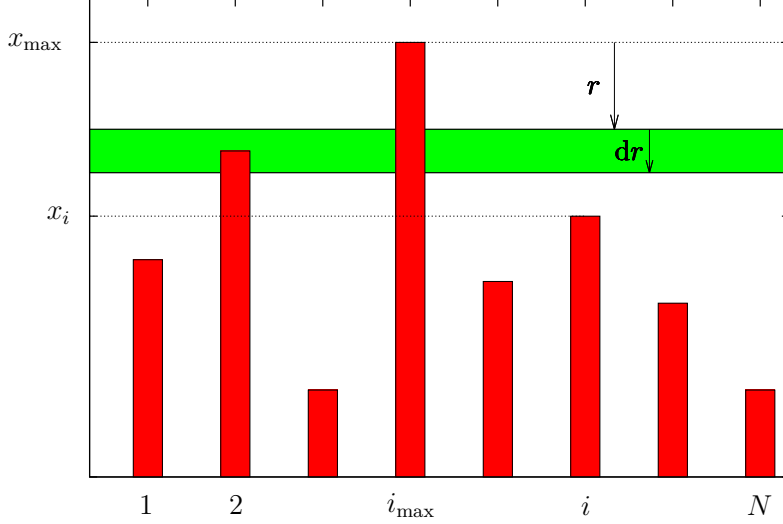


FIGURE 4: Pour une collection de variables $\{x_i | i = 1, \dots, N\}$, la densité à partir du maximum $\rho_{\max}(r, N)dr$ compte le nombre de variables qui sont à une distance x du maximum à dr près représenté par la bande verte sur la figure.

$\langle \rho_0(x, N) \rangle$ caractérise alors le nombre moyen d'états d'énergie $E = -x$ où la moyenne $\langle \dots \rangle$ est prise suivant la loi des variables aléatoires. En suivant cette quantité, on introduit la densité à partir du maximum

$$\rho_{\max}(r, N) = \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq i_{\max}}}^N \delta(x_{\max} - r - x_i), \quad (5)$$

qui compte les états situés en $x_{\max} - r$ c'est-à-dire à une distance r du maximum qui est lui-même une variable aléatoire. En effet, $\rho_{\max}(r, N)dr$ est le nombre de variables comprises entre $x_{\max} - r - dr$ et $x_{\max} - r$ comme le montre la figure 4. On a noté dans la définition (5) i_{\max} l'indice tel que $x_{\max} = x_{i_{\max}}$ de manière à avoir une densité normalisée $\int_0^\infty dr \rho_{\max}(r, N) = 1$. Il est également possible d'étudier cette quantité à partir du minimum. Une autre quantité intéressante est la statistique du premier gap, qui dans la correspondance avec la physique statistique correspond à l'énergie du premier état excité. Si on renumérote les variables aléatoires par ordre décroissant $x_{\max} = X_1 > X_2 > X_3 > \dots > X_N$, le premier gap est donné par $d_1 = X_1 - X_2$

$$p_{\text{gap}}(x, N) = \langle \delta(X_1 - X_2 - x) \rangle. \quad (6)$$

Ainsi $p_{\text{gap}}(x, N)dx$ est la probabilité que le gap $X_1 - X_2$ entre les deux plus grandes variables soit x à dx près. Il est également possible d'étudier le k -ième gap $d_k = X_k - X_{k+1}$. Ces quantités ont été largement analysées dans le cas des variables indépendantes et identiquement distribuées [35, 77, 139], il existe assez peu de résultats pour le cas des variables aléatoires fortement corrélées [21, 22, 68, 70, 107, 115, 142, 166]. Le but de cette thèse est ainsi de faire ce premier pas en étudiant la statistique de la densité à partir du maximum de deux modèles de variables aléatoires corrélées pour lesquels il est possible de tirer des résultats analytiques exacts : le mouvement brownien et les matrices aléatoires.

Organisation du manuscrit

Cette thèse est ainsi consacrée à l'étude de statistique d'extrêmes pour les deux ensembles de variables fortement corrélées que sont les positions d'un mouvement brownien et les valeurs propres de matrices aléatoires. Je rappelle en premier lieu dans le chapitre 1 les résultats pour des variables indépendantes et identiquement distribuées. Ce cas a été beaucoup étudié par les mathématiciens et physiciens. Il a été montré que la distribution du maximum p_{\max} dans la limite N vers l'infini, après renormalisation et centrage, ne peut tendre que vers trois lois. Ce résultat permet alors d'étudier dans ce cas les autres quantités intéressantes que sont la densité à partir du maximum et le gap.

Dans la première partie, j'étudie le cas où la collection de variables aléatoires est la position $x(\tau)$ au cours du temps $\tau \in [0, t]$ d'un mouvement brownien, qui peut être contraint à être périodique, positif... Ce mouvement brownien est vu comme la limite d'un marcheur aléatoire classique qui fait un grand nombre de pas. On réinterprètera alors ce problème comme celui d'une particule quantique dans un potentiel ce qui permet d'utiliser des méthodes puissantes issues de la mécanique quantique comme l'utilisation de propagateurs (chapitre 3) et de l'intégrale de chemin (chapitre 5). Ces outils permettent de calculer la densité moyenne à partir du maximum pour les différents mouvements browniens contraints (3.18, 3.28, 3.29, 3.30) et même la distribution complète de cette quantité pour certains cas (3.25, 5.35, 5.36, 5.61). Nous généraliserons notre démarche à l'étude de plusieurs marches aléatoires indépendantes ou avec interaction ainsi qu'à l'étude temporelle des positions à une distance r du maximum dans le chapitre 4. Nous appliquerons également nos résultats pour calculer la statistique d'une certaine famille de fonctionnelle du maximum (5.69) qui possède de multiples applications que nous discuterons. Les principaux résultats de cette partie ont été publiés dans deux articles : le premier article, paru dans Physical review letters [122], traite de la statistique de la densité à partir des extrêmes pour les différents mouvements browniens tandis que le second, en version preprint [123], étudie plus généralement les fonctionnelles du maximum et leurs applications en particulier à l'algorithme de recherche en environnement aléatoire que nous développons dans l'annexe E. Nous retrouvons en particulier d'une autre manière les résultats de Chassaing *et al.* [28].

Dans la seconde partie, j'étudie le cas où la collection de variables aléatoires est composée des N valeurs propres $\{\lambda_i | i = 1, \dots, N\}$ d'une matrice aléatoire $N \times N$. Ces matrices aléatoires seront générées suivant les modèles courants que sont les matrices carrées gaussiennes, matrices dont les entrées sont des gaussiennes indépendantes, ou les matrices de Wishart, matrices de covariance pour des données statistiques. Nous étudierons pour les deux modèles le voisinage de la valeur propre maximale grâce à une méthode fondée sur les propriétés des polynômes orthogonaux. Dans le cas des matrices gaussiennes unitaires, nous obtiendrons une formule analytique (7.26) pour la distribution à partir du maximum dont nous étudierons les différentes limites d'échelles (7.29) et (7.43). Nous obtenons également une nouvelle expression de la statistique du gap entre les deux plus grandes valeurs propres (7.47) qui avait été étudié par Witte *et al.* [166]. Ces différents résultats ont été publiés dans la revue Journal of Statistical Physics [126]. Nous appliquerons également nos résultats, et plus particulièrement leurs généralisations aux différents cas gaussien, développées dans la section 8.1, à un modèle de verre de spin en champs moyen dans la section 8.3. Nous analyserons également dans la section 8.4 la densité à partir du maximum, ainsi que le premier gap, lorsque la valeur propre maximale est contrainte. Ces différentes généralisations sont traitées dans la prépublication [127]. L'application au modèle de verre spin a donné lieu à une collaboration avec Y.V.Fyodorov aboutissant à un papier à paraître [124]. Pour le cas des matrices de Wishart, nous redérivons dans le

chapitre 9 en utilisant les polynômes orthogonaux les différentes statistiques de la valeur propre minimale dont nous montrons dans la section 9.3 qu'elle est gouvernée par une équation de Painlevé V à N fini, dans la section 9.4 qu'elle est gouvernée par une équation de Painlevé III dans la limite dite de *hard edge* et dans la section 9.5 qu'elle est gouvernée par une équation de Painlevé II - comme pour le cas gaussien - dans la limite dite de *soft edge*. Nous prouvons également une conjecture faite par Edelman *et al.* [47] sur la première correction de la distribution de la valeur propre minimale dans la limite de *hard edge* (9.55). Cette redérivation des équations de Painlevé pour les différents régimes sera traitée dans un article à paraître [128]. Nous terminerons par étudier intégralement le cas des matrices de Wishart $a = 0$ pour lesquelles nous analyserons la densité proche du minimum dans l'appendice G.

Notations et formulaires

Nous noterons $\langle X \rangle$ la moyenne suivant les variables aléatoires considérées qui dépendra du contexte : moyenne sur les trajectoires browniennes (pouvant être contraintes) dans la première partie et moyenne sur la distribution des valeurs propres des matrices aléatoires analysées dans la seconde.

Dans la première partie, nous noterons $x(\tau)$ une trajectoire de brownien libre sur $[0, t]$ c'est-à-dire un mouvement brownien démarrant de la position $x(0) = 0$ et terminant en une position finale $x(t) = x_F$ libre. Nous noterons $M = x_{\max}$ la position maximale de cette trajectoire qui est atteinte en un temps que nous noterons t_{\max}

$$x_{\max} = \max_{\tau \in [0, t]} x(\tau) = x(t_{\max}). \quad (7)$$

Nous utiliserons les indices :

- B lorsque l'on contraint la trajectoire à être un pont, c'est-à-dire à terminer en $x_F = 0$.
- E lorsque l'on contraint la trajectoire à être une excursion, c'est-à-dire à terminer en $x_F = 0$ et à rester positive tout du long $x(\tau) \geq 0$ pour $\tau \in [0, t]$.
- Me lorsque l'on contraint la trajectoire à être un méandre, c'est-à-dire à rester positive tout du long $x(\tau) \geq 0$ pour $\tau \in [0, t]$.

Nous utiliserons également l'exposant R lorsque nous considérons des trajectoires réfléchies c'est-à-dire la valeur absolue d'une trajectoire.

Nous noterons $\tilde{f}(s)$ la transformée de Laplace d'une fonction $f(t)$ par rapport à son argument définie par

$$\tilde{f}(s) = \int_0^\infty dt f(t) e^{-st}. \quad (8)$$

Nous noterons ∂_x la dérivée par rapport à un paramètre x .

Dans la seconde partie de ce manuscrit, nous rencontrerons souvent la fonction $q(x)$ de Hastings-McLeod qui est la solution de l'équation de Painlevé II vérifiant

$$\begin{cases} q''(x) = 2q^3(x) + xq(x), \\ q(x) \sim \text{Ai}(x) \text{ pour } x \rightarrow \infty, \end{cases} \quad (9)$$

ainsi que la distribution de Tracy-Widom \mathcal{F}_2

$$\mathcal{F}_2(x) = \exp \left(- \int_x^\infty du (u - x) q^2(u) \right). \quad (10)$$

On rencontrera au cours de ce manuscrit les fonctions spéciales courantes suivantes :

- La fonction d'erreur et son complémentaire

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = 1 - \text{erfc}(x). \quad (11)$$

- La fonction d'Airy

$$\text{Ai}(x) = \frac{1}{\pi} \int_0^\infty \cos \left(\frac{t^3}{3} + xt \right) dt. \quad (12)$$

- La fonction Gamma

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad (13)$$

en particulier $\Gamma(n+1) = n!$ pour n entier.

— La fonction bêta incomplète

$$B_z(a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt. \quad (14)$$

— La fonction de Bessel de première espèce J_ν ainsi que la fonction de Bessel modifiée de première espèce I_ν

$$J_\nu(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \nu + 1)} \left(\frac{x}{2}\right)^{2n+\nu}, \quad (15)$$

$$I_\nu(x) = \sum_{n=0}^{\infty} \frac{1}{n! \Gamma(n + \nu + 1)} \left(\frac{x}{2}\right)^{2n+\nu}. \quad (16)$$

— La fonction hypergéométrique confluyente de Kummer

$${}_1F_1(a, b, z) = \sum_{n=0}^{\infty} \frac{\Gamma(b) \Gamma(a+n)}{\Gamma(b+n) \Gamma(a)} \frac{z^n}{n!}. \quad (17)$$

— La fonction hypergéométrique confluyente de Tricomi $b \notin \mathbb{Z}$

$$U(a, b, z) = \frac{\Gamma(1-b)}{\Gamma(a-b+1)} {}_1F_1(a, b, z) + \frac{\Gamma(b-1)}{\Gamma(a)} z^{1-b} {}_1F_1(a-b+1, 2-b, z). \quad (18)$$

— La fonction êta de Dirichlet ainsi que la fonction zêta de Riemann

$$\eta(s) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^s} = (1 - 2^{1-s}) \zeta(s), \quad \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \quad (19)$$

Nous utiliserons également les familles de polynômes :

— d'Hermite

$$H_n(x) = n! \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{(-1)^i (2x)^{n-2i}}{i! (n-2i)!}, \quad (20)$$

avec $\lfloor x \rfloor$ la partie entière de x , vérifiant la condition de normalisation

$$\int_{-\infty}^{\infty} dx e^{-x^2} H_n(x) H_m(x) = 2^n n! \sqrt{\pi} \delta_{n,m}. \quad (21)$$

— de Laguerre généralisés

$$L_n^{(a)}(x) = \frac{\Gamma(a+n+1)}{n!} \sum_{k=0}^n \binom{n}{k} \frac{(-x)^k}{\Gamma(a+k+1)}, \quad (22)$$

vérifiant la condition de normalisation

$$\int_0^{\infty} dx e^{-x} x^a L_n^{(a)}(x) L_m^{(a)}(x) = \frac{\Gamma(n+a+1)}{n!} \delta_{n,m}. \quad (23)$$

Chapitre 1

Statistique d'extrêmes de variables aléatoires indépendantes

Le cas des variables aléatoires indépendantes et identiquement distribuées est bien compris grâce aux différents travaux [61, 71, 76, 77, 78, 139]. On connaît désormais principalement tout de la statistique d'extrêmes dans ce cas. Je rappelle dans ce chapitre les principaux résultats. Dans la section 1.1, nous étudierons la statistique du maximum qui converge dans ce cas, après renormalisation et centrage, uniquement vers 3 lois que sont les lois de Gumbel, de Fréchet et de Weibull. Dans la section suivante 1.2, nous discuterons de la densité à partir du maximum où également plusieurs cas se présentent. Dans la dernière section 1.3, nous terminerons par mentionner quelques résultats sur la statistique d'ordre dans ce cas et plus particulièrement sur le premier gap.

On se donne une collection de N variables réelles $\{x_i | i = 1, \dots, N\}$ indépendantes et de même densité de probabilité $p(x)$, c'est-à-dire que la loi jointe des N variables se factorise $P_{\text{joint}}(x_1, x_2, \dots, x_N) = p(x_1)p(x_2) \cdots p(x_N)$. On note par la suite X^* la borne supérieure du support de p (qui peut être infinie)

$$X^* = \sup_x \{p(x) > 0\}. \quad (1.1)$$

Il est clair que dans la limite de grand N , le maximum des variables tend vers cette borne X^* .

1.1 Statistique du maximum

La théorie des statistiques d'extrêmes s'intéresse en particulier à la distribution de la plus grande variable $x_{\text{max}} = \max\{x_i | i = 1, \dots, N\}$

$$p_{\text{max}}(x, N) = \text{Pr}.[x_{\text{max}} < x]. \quad (1.2)$$

Il est possible de donner une estimation μ_N du maximum pour une collection de N variables indépendantes et de même distribution

$$\int_{\mu_N}^{X^*} dx p(x) = \frac{1}{N}, \quad (1.3)$$

qui stipule qu'il n'y a en moyenne qu'une et une seule des N variables - le maximum x_{max} - supérieure à μ_N . De cette formule, il est facile de montrer que dans la limite thermodynamique d'un grand nombre de variables, $N \rightarrow \infty$, le maximum de la collection et donc μ_N tend vers X^* . Comme les variables sont indépendantes, il est possible de calculer

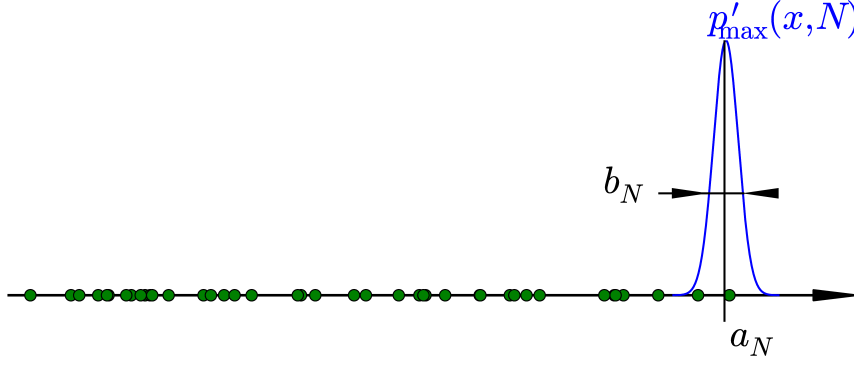


FIGURE 1.1: La distribution du maximum $p'_{\max}(x, N)dx = \Pr.[x_{\max} \in [x, x + dx]]$ est centrée autour de a_N et a pour extension typique b_N . Il est nécessaire de renormaliser par ces deux paramètres pour trouver une distribution limite non triviale.

analytiquement la fonction de répartition du maximum $p_{\max}(x, N)$ en contraignant les N variables aléatoires à être plus petites que x

$$p_{\max}(x, N) = \Pr.[x_{\max} < x] = \left(\int_{-\infty}^x dz p(z) \right)^N. \quad (1.4)$$

Prenons pour exemple le cas de N variables exponentielles indépendantes $p(x) = e^{-x}$ si $x \geq 0$ et 0 sinon. Dans ce cas, il est facile de calculer la fonction de répartition du maximum (1.4)

$$p_{\max}(x, N) = (1 - e^{-x})^N, \quad (1.5)$$

que l'on peut réécrire

$$p_{\max}(x, N) = \left(1 - \frac{1}{N} e^{-(x - \ln N)} \right)^N \underset{N \rightarrow \infty}{\sim} e^{-e^{-(x - \ln N)}}. \quad (1.6)$$

Ainsi, dans la limite N tendant vers l'infini,

$$\Pr.[x_{\max} - \ln N < x] \underset{N \rightarrow \infty}{\rightarrow} e^{-e^{-x}}. \quad (1.7)$$

Dans le cas général, on peut montrer que dans la limite où N est grand, à un recentrage donné par un paramètre a_N et une normalisation donnée par un paramètre $b_N > 0$ pour avoir une limite non triviale, comme le montre la figure 1.1, il n'existe que trois possibilités tracées en figure 1.2 pour la densité de probabilité limite suivant le comportement de p au voisinage de X^* [61, 71, 76]. On parle des classes d'universalités de Gumbel, de Fréchet et de Weibull.

1.1.1 la classe de Gumbel

Si la borne supérieure du support X^* est infinie et si $p(x)$ décroît plus rapidement que toute puissance à grand x , il s'agit par exemple des cas des variables exponentielles et gaussiennes (figure 3(c)), alors la loi limite est la loi de Gumbel. En prenant $a_N = \mu_N$ défini en (1.3) et b_N l'écart typique entre le maximum de la collection x_{\max} et son estimation a_N donnée par

$$b_N = \frac{\int_{a_N}^{X^*} dx (x - a_N) p(x)}{\int_{a_N}^{X^*} dx p(x)}, \quad (1.8)$$

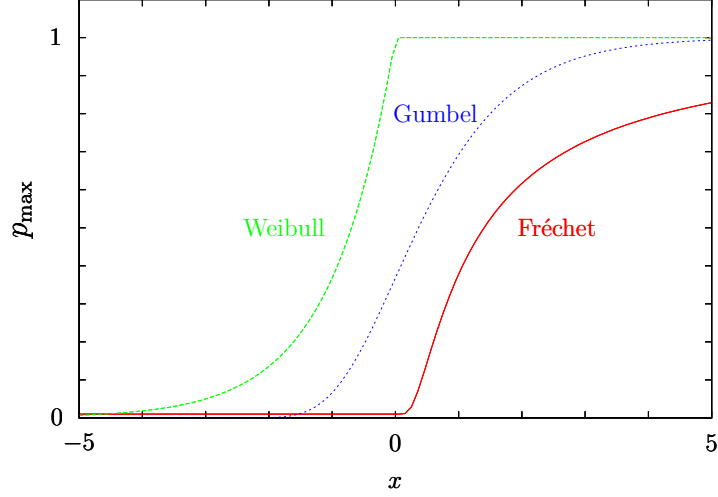


FIGURE 1.2: Les trois lois limites possibles pour la densité de probabilité limite du maximum recentrée et renormalisée dans le cas de variables indépendantes : la loi de Fréchet $\gamma = 1$, la loi de Gumbel et la loi de Weibull $\gamma = 1$.

où l'on a utilisé la probabilité conditionnée à n'avoir qu'une seule variable, le maximum x_{\max} , plus grande que a_N , alors la distribution du maximum renormalisée tend vers une loi de Gumbel

$$\Pr. \left[\frac{x_{\max} - a_N}{b_N} < x \right] \xrightarrow{N \rightarrow \infty} G_I(x) = e^{-e^{-x}}. \quad (1.9)$$

C'est le résultat que nous avons obtenu pour l'exemple des variables exponentielles ci-dessus avec $a_N = \ln N$ et $b_N = 1$. Notons qu'il existe aussi des cas pathologiques où la borne supérieure du support X^* est finie mais qui sont tout de même dans la classe d'universalité de Gumbel, il s'agit par exemple des cas où la loi possède une singularité essentielle en X^* comme $p(x) \propto e^{-1/(X^*-x)^\nu}$ avec $\nu > 0$.

1.1.2 la classe de Fréchet

Si la borne supérieure du support X^* est infinie et si $p(x)$ décroît comme une loi de puissance $x^{-1-\gamma}$ à grand x , il s'agit par exemple des cas des variables de Cauchy (figure 3(b)), de Student et de Pareto, alors la loi limite est la loi de Fréchet. En prenant $a_N = 0$ et $b_N = \mu_N$ (1.3), alors la distribution du maximum renormalisée tend vers une loi de Fréchet

$$\Pr. \left[\frac{x_{\max} - a_N}{b_N} < x \right] \xrightarrow{N \rightarrow \infty} G_{II}(x) = \begin{cases} 0 & x < 0 \\ e^{-x^{-\gamma}} & x \geq 0 \end{cases}. \quad (1.10)$$

1.1.3 la classe de Weibull

Si la borne supérieure du support X^* est finie et que le comportement de $p(x)$ au voisinage de X^* est de type $p(x) \propto (X^* - x)^{\gamma-1}$, il s'agit par exemple des cas des variables uniforme sur un segment (figure 3(a)), des lois beta ainsi que de la loi du demi-cercle de Wigner que nous verrons lors de l'étude des matrices aléatoires, alors la loi limite est la loi de Weibull. En prenant $a_N = X^*$ et $b_N = X^* - \mu_N$ (1.3), alors la distribution du

maximum renormalisée tend vers une loi de Weibull

$$\Pr. \left[\frac{x_{\max} - a_N}{b_N} < x \right] \xrightarrow{N \rightarrow \infty} G_{\text{III}}(x) = \begin{cases} e^{-|x|^\gamma} & x < 0 \\ 1 & x \geq 0 \end{cases} . \quad (1.11)$$

1.1.4 Remarques

Ainsi, dans chacun des cas, il existe des coefficients a_N et b_N tels que

$$p_{\max}(b_N x + a_N, N) = \Pr. \left[\frac{x_{\max} - a_N}{b_N} < x \right] \xrightarrow{N \rightarrow \infty} G_k(x), \quad (1.12)$$

où $k = \text{I}$ dans le bassin d'attraction de la loi de Gumbel (1.9), $k = \text{II}$ dans le bassin d'attraction de la loi de Fréchet (1.10) et $k = \text{III}$ dans le bassin d'attraction de la loi de Weibull (1.11). Il y a donc trois cas suivant le comportement de la distribution $p(x)$ à grand argument. Ces résultats décrivent la statistique du maximum dans la limite de grand N mais ne donne aucune indication sur la distribution à N fini. Dans ce cas, les effets de tailles finies ont été également largement étudiés [13, 37, 38, 61, 79, 80] et décroissent très lentement en $\mathcal{O}(\log(N))$. Le comportement du coefficient b_N dans la limite de grand N dépend également de la queue de $p(x)$ à grand x

$$\lim_{N \rightarrow \infty} b_N = \begin{cases} 0 & \text{si } p(x) \text{ décroît plus vite qu'une exponentielle} \\ \text{constante} & \text{si } p(x) \text{ à un comportement exponentiel} \\ \infty & \text{si } p(x) \text{ décroît moins vite qu'une exponentielle.} \end{cases} \quad (1.13)$$

Comme le montre la figure 1.1, la distribution du maximum aura tendance à s'étaler si la queue de $p(x)$ décroît moins vite qu'une exponentielle alors qu'au contraire elle aura tendance à devenir de plus en plus piquée si la queue de $p(x)$ décroît plus vite qu'une exponentielle jusqu'à conduire à un pic $\delta(x - a_N)$.

La distribution du maximum dans le cas de variables aléatoires indépendantes et identiquement distribuées et ainsi très bien comprise. Néanmoins, l'étude de cette statistique n'apporte qu'une information partielle. En particulier, il ne répond pas à la question de savoir si ce maximum est proche d'autres variables aléatoires ou si au contraire il est loin de toutes les autres variables. Pour cela, nous étudions la densité à partir du maximum (5).

1.2 Densité moyenne à partir du maximum

En utilisant encore une fois l'indépendance des variables, il est possible de réécrire la densité moyenne à partir du maximum (5) comme [139]

$$\langle \rho_{\max}(r, N) \rangle = \frac{N}{N-1} \int_{-\infty}^{+\infty} \overbrace{dz}^{\text{valeur du maximum}} \underbrace{p(z-r)}_{\text{une variable à } x \text{ du maximum}} \overbrace{p'_{\max}(z, N-1)}^{\text{le maximum des } N-1 \text{ autres variables est } z} . \quad (1.14)$$

On peut interpréter cette formule : la densité moyenne $\langle \rho_{\max}(r, N) \rangle$ compte le nombre de variables aléatoires à une distance r de la variable maximale. Cette quantité est la même pour chacune des N variables, on peut donc obliger x_N à être à une distance r du maximum des autres variables $\max\{x_i, i = 1, \dots, N-1\}$ et multiplier le résultat par N . Ce maximum est z à dz près avec une probabilité $\Pr. [\max\{x_i, i = 1, \dots, N-1\} \in [z, z + dz]] = p'_{\max}(z, N-1)dz$ et x_N est en $z - r$ avec une probabilité $p(z - r)$. Il ne reste alors qu'à intégrer sur toutes les valeurs de z possibles pour trouver le résultat.

Reprenons l'exemple des N variables exponentielles indépendantes, dans ce cas, la formule (1.14) donne directement

$$\langle \rho_{\max}(r, N) \rangle = \frac{e^r}{(n-1)} \left(1 - \left(1 + (n-1)e^{-r} \right) \left(1 - e^{-r} \right)^{n-1} \right). \quad (1.15)$$

Ainsi, en reprenant la même limite que précédemment (1.7), la densité à partir du maximum dans le cas exponentiel prend la forme limite

$$\langle \rho_{\max}(z + \log N, N) \rangle \xrightarrow{N \rightarrow \infty} e^z \left(1 - \left(1 + e^{-z} \right) e^{-e^{-z}} \right). \quad (1.16)$$

Il est possible à l'aide des résultats précédents d'étudier plus précisément la distribution limite de $\langle \rho_{\max}(r, N) \rangle$ pour n'importe quelle distribution $p(x)$ [139]. Le comportement de cette distribution dépend de la comparaison de la queue à grand r de la distribution $p(x)$ avec une exponentielle. Le cas exponentiel ayant été traité précédemment (1.16), il ne reste qu'à explorer les deux autres cas possibles :

- Si la loi décroît moins vite qu'une exponentielle, alors le coefficient b_N qui caractérise l'étalement de la distribution du maximum tend vers l'infini. En utilisant le changement de variable $z \rightarrow b_N y + a_N$ dans l'équation (1.14), on trouve

$$\langle \rho_{\max}(r, N) \rangle = \frac{N}{N-1} \int_{-\infty}^{+\infty} b_N dy p(b_N y + a_N - r) p'_{\max}(b_N y + a_N, N-1). \quad (1.17)$$

Or, puisque b_N tend vers l'infini, $b_N p(b_N y + a_N - r)$ est une distribution de plus en plus piquée autour de la valeur $y = \frac{r - a_N}{b_N}$, elle tend donc vers le pic de Dirac $\delta\left(y - \frac{r - a_N}{b_N}\right)$. En utilisant le résultat précédent, la distribution du maximum $b_N p'_{\max}(b_N y + a_N, N-1)$ tend vers $G'_k(z)$ suivant le bassin d'attraction dans lequel se trouve la distribution $p(x)$ (Gumbel ou Fréchet). Finalement, la densité à partir du maximum prend la forme d'échelle limite

$$\lim_{N \rightarrow \infty} b_N \langle \rho_{\max}(b_N z + a_N, N) \rangle = G'_k(z), \quad (1.18)$$

où $k = \text{I}$ dans le bassin d'attraction de la loi de Gumbel (1.9), $k = \text{II}$ dans le bassin d'attraction de la loi de Fréchet (1.10).

- Si la loi décroît plus vite qu'une exponentielle, alors le coefficient b_N tend vers 0. Dans ce cas, la distribution du maximum $p'_{\max}(z, N-1)$ devient de plus en plus piquée jusqu'à conduire à un pic $\delta(x - a_N)$, les fluctuations du maximum d'ordre b_N devenant négligeable. En utilisant la formule (1.14), on trouve directement le comportement limite de la densité à partir du maximum

$$\lim_{N \rightarrow \infty} \langle \rho_{\max}(a_N - z, N) \rangle = p(z). \quad (1.19)$$

Remarquons le rôle particulier que joue la distribution exponentielle qui délimite deux régions de comportements bien différents [121]. La quantité $\langle \rho_{\max}(0, N) \rangle$ est le nombre moyen de variables proches du maximum, plus précisément $\langle \rho_{\max}(0, N) \rangle dx$ est le nombre moyen de variables à moins de dx du maximum, elle peut parfois être obtenue à partir des comportements limites précédents avec $r = 0$ si celui-ci appartient à la loi d'échelle. Sinon, il est toujours possible de calculer $\langle \rho_{\max}(0, N) \rangle$ directement à partir de la formule (1.14) ou, comme nous allons le voir, en utilisant les résultats sur le premier gap.

1.3 Premier gap et statistiques d'ordre

Une autre quantité intéressante permettant d'étudier les variables proches du maximum est la distribution du premier gap et plus généralement les distributions du i -ième gap. En renumérotant les N variables x_1, \dots, x_N par ordre décroissant $X_1 \geq \dots \geq X_N$, le i -ième gap est défini par $g_i = X_i - X_{i+1}$. Les distributions des différents maxima ainsi que des gaps pour des variables indépendantes ont été largement étudiées [4, 12, 35, 143]. De la même façon que le maximum, il est nécessaire de renormaliser la variable étudiée pour obtenir une distribution limite non triviale, cette renormalisation est réalisée par la même loi d'échelle pour le i -ième maximum (i fini)

$$\lim_{N \rightarrow \infty} \frac{X_i - a_N}{b_N} = W_i. \quad (1.20)$$

Dans ce cas, la fonction de répartition limite du i -ième maximum est donnée par [35, 143]

$$\begin{aligned} \lim_{N \rightarrow \infty} \Pr. \left[W_i = \frac{X_i - a_N}{b_N} < x \right] &= G_k(x) \sum_{j=0}^{i-1} \frac{(-\ln G_k(x))^j}{j!} \\ &= \frac{1}{(i-1)!} \int_{-\ln G_k(x)}^{\infty} dt e^{-t} t^{i-1}, \end{aligned} \quad (1.21)$$

qui redonne bien le résultat (1.12) pour le premier maximum $X_1 = x_{\max}$. Il est également possible d'étudier la distribution jointe des i premiers maxima, celle-ci est donnée par

$$P_{\text{joint}}(W_1 > W_2 > \dots > W_i) = G_k(W_i) \prod_{j=1}^i \frac{G'_k(W_j)}{G_k(W_j)}. \quad (1.22)$$

La renormalisation nécessaire sur le i -ième maximum (1.20) détermine la loi d'échelle de la distribution du i -ième gap $p_i(r, N)dr = \Pr.[g_i = X_i - X_{i+1} \in [r, r + dr]]$ avec $r > 0$. Il est possible de donner une fonction analytique pour chacun des bassins d'attractions précédents

$$\lim_{N \rightarrow \infty} b_N p_i(b_N r, N) = p_{\text{gap}, i}(r) \quad (1.23)$$

$$= \frac{1}{(i-1)!} \int_{-\infty}^{\infty} dx G'_k(x) \frac{G'_k(x+r)}{G_k(x+r)} [-\ln G_k(x+r)]^{i-1}. \quad (1.24)$$

Appliqué au trois fonctions limites possibles G_k , on trouve finalement

$$p_{\text{gap}, i}(r) = \begin{cases} i e^{-ir} & \text{Gumbel} \\ \frac{\gamma^2}{(i-1)!} \int_0^{\infty} dx e^{-x^{-\gamma}} x^{-\gamma-1} (x+r)^{-\gamma i-1} & \text{Fréchet} \\ \frac{\gamma^2}{(i-1)!} \int_0^{\infty} dx (x+r)^{\gamma-1} e^{-(x+r)^{\gamma}} x^{\gamma i-1} & \text{Weibull.} \end{cases} \quad (1.25)$$

Remarquons qu'il est possible d'adapter la formule (1.14) pour calculer la distribution du premier gap, écart entre les deux plus grandes variables X_1 et X_2 en utilisant encore une fois l'indépendance des variables

$$p_1(r, N) = N \int_{-\infty}^{+\infty} \underbrace{dz}_{\text{valeur de } X_2} \underbrace{p(z+r)}_{X_1 \text{ est en } X_2 + r} \overbrace{p'_{\max}(z, N-1)}^{\text{le maximum des variables autres que } X_1 \text{ est } z = X_2} = (N-1) \langle \rho_{\max}(-r, N) \rangle. \quad (1.26)$$

Le premier gap est donc relié, dans ce cas, à la densité proche du maximum par la transformation $r \rightarrow -r$ qui revient à considérer cette densité en dehors de son domaine

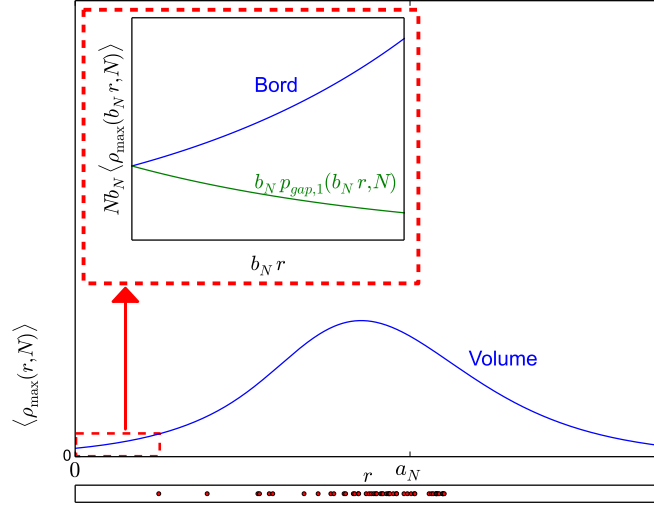


FIGURE 1.3: La densité à partir du maximum possède plusieurs régimes d'échelle. Le régime dit de volume considère des variables à une distance typique a_N du maximum, ce qui est le cas de la majorité des variables. Il s'agit du régime étudié dans la section 1.2 [139]. Le régime de bord s'obtient en considérant des r d'ordre b_N^{-1} . Dans ce cas, on considère les premières variables les plus proches du maximum. Ce régime fait le lien entre le premier gap (à r petit) et la queue gauche du régime de bord (à r grand). La figure a été tracée en utilisant les lois théoriques du cas exponentiel avec un dessous un exemple numérique de 50 écarts $x_{\max} - x_i$ obtenus.

naturel de définition $r > 0$. Il n'est pas forcément possible d'obtenir le régime typique du premier gap (1.23) à partir de la loi d'échelle de la densité proche du maximum obtenue dans la section 1.2. Pour employer un vocabulaire que nous redéfinirons plus précisément dans la partie II sur les matrices aléatoires, le régime de la densité à partir du maximum étudié dans la section 1.2 est le régime de volume (**Bulk**). Dans ce régime, on étudie les variables qui sont à une distance de l'ordre de a_N du maximum ce qui est le cas de la majorité des variables. Le régime typique du premier gap (1.23) est le régime de bord (**Edge**) pour lequel on étudie seulement les quelques premières variables proches du maximum. Il est possible d'observer le régime de bord de la densité à partir du maximum $Nb_N \langle \rho_{\max}(b_N r, N) \rangle$. En particulier, dans la limite $r = 0$, on a

$$\lim_{N \rightarrow \infty} Nb_N \langle \rho_{\max}(0, N) \rangle = p_{gap,1}(0). \quad (1.27)$$

Ainsi, la densité très proche du maximum peut directement être calculée en utilisant les comportements du premier gap (1.25). La limite à r grand de $Nb_N \langle \rho_{\max}(b_N r, N) \rangle$ doit venir quant à elle se raccorder avec le comportement de volume obtenu dans la section 1.2. Par exemple, en reprenant le cas exponentiel (1.15) pour lequel $b_N = 1$, le comportement de bord de la densité à partir du maximum est donné par

$$\lim_{N \rightarrow \infty} N \langle \rho_{\max}(r, N) \rangle = e^r. \quad (1.28)$$

Dans ce cas, on retrouve $p_{gap,1}(-r)$ donné par le cas Gumbel dans (1.25) et donc la limite de $r = 0$ (1.27) est vérifié. Le comportement (1.28) vient bien se raccorder avec la loi limite (1.16) dans la limite z vers $-\infty$. Le schéma de la figure 1.3 représente les différentes échelles considérées.

Première partie
Mouvement brownien

Chapitre 2

Introduction

Après quelques brefs rappels sur le mouvement brownien, nous introduisons dans la section 2.2 les différents mouvements browniens contraints que nous serons amenés à étudier. Nous rappellerons alors quelques résultats sur la statistique du maximum pour les différents mouvements browniens dans la section 2.3 et définissons dans la section 2.4 la densité à partir du maximum pour des mouvements browniens.

2.1 Le mouvement brownien

2.1.1 Présentation

Le mouvement brownien, introduit en 1905 par Einstein [50] et Von Smoluchowski [162] pour expliquer le mouvement de particules dans des grains de pollen [18] est l'un des domaines les plus prolifiques de la physique avec des applications diverses dans toutes les branches des sciences : mathématiques [57], physique [159], biologie [10]... Il est néanmoins impossible d'en faire ici une liste exhaustive, il existe une multitude de revues publiées pour les 100 ans de son introduction comme par exemple [29, 45, 72, 93, 103]. Le mouvement brownien, bien que largement étudié à travers cette abondance d'articles, reste toujours un modèle moderne que l'on peut analyser à l'aide de méthodes puissantes.

2.1.2 Définition

Le mouvement brownien (aussi appelé processus de Wiener) sur un intervalle de temps $[0, t]$ est un processus stochastique continu $x(\tau)$ avec $\tau \in [0, t]$:

- qui possède la propriété de Markov c'est-à-dire une absence de mémoire. Le processus pour $\tau \in [u, t]$ ne dépend que de la valeur $x(u)$ et pas de la trajectoire complète sur $[0, u]$
- dont le propagateur $G(\alpha|\beta, t)$ des trajectoires qui démarrent en $x(0) = \alpha$ et terminent après un temps t en $x(t) = \beta$ est une variable aléatoire gaussienne stationnaire

$$G(\alpha|\beta, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(\alpha-\beta)^2}{2t}}. \quad (2.1)$$

Le mouvement brownien est invariant d'échelle, c'est-à-dire que ces propriétés sont similaires peu importe à quelle échelle on se place. Plus précisément, si l'on allonge l'échelle de distances $x \rightarrow \alpha x$ et étend l'échelle de temps $t \rightarrow \alpha^2 t$, les propriétés du système ne changent pas, la bonne variable à considérer est ainsi x/\sqrt{t} . On appellera par la suite un brownien libre un tel processus considéré sur un intervalle de temps $[0, t]$ dont la position

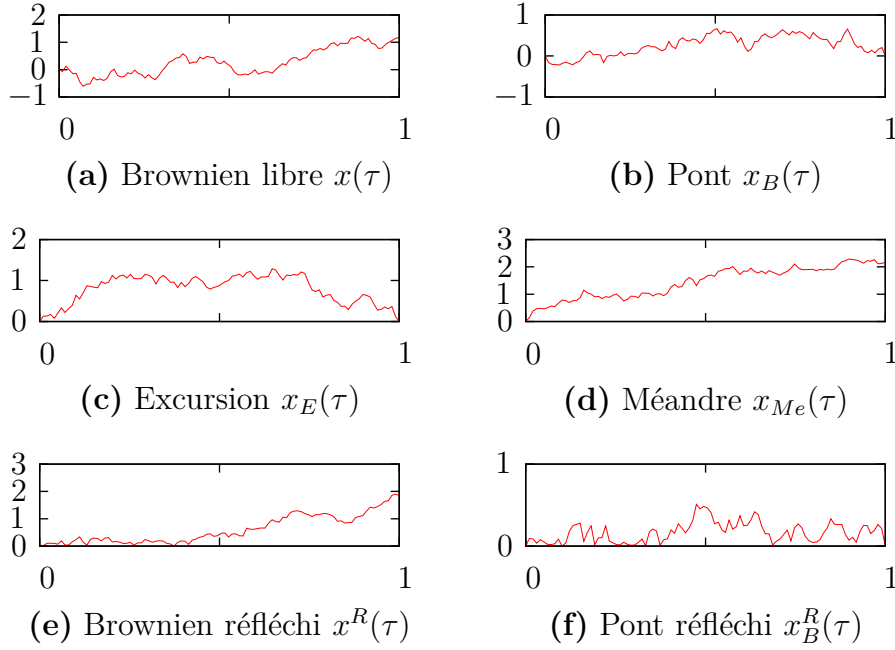


FIGURE 2.1: Simulation des différents browniens contraints considérés à partir de marches aléatoires de $N = 100$ pas. **(a)** Le brownien libre est généré à partir de la marche (2.2). **(b)** Le pont, brownien contraint à revenir à l'origine, est généré grâce à (2.3). **(c)** L'excursion, brownien positif contraint à revenir à l'origine, est généré à partir de 3 réalisations de la marche (2.2) grâce à la propriété (2.4). **(d)** L'excursion, brownien positif, est généré à partir de 3 réalisations de la marche (2.2) via la propriété (2.5). **(e)** et **(f)** Le brownien réfléchi (respectivement le pont réfléchi), est la valeur absolue d'une trajectoire de brownien libre (respectivement de pont).

initiale est fixée à l'origine $x(0) = 0$ et l'autre extrémité $x(t)$ reste libre. Un brownien libre sur $[0, t]$ peut être vu comme la limite de la marche aléatoire discrète

$$\begin{cases} X_0 = 0 \\ X_i = X_{i-1} + \eta_i, i \in [1, N], \end{cases} \quad (2.2)$$

où les η_i sont des variables gaussiennes indépendantes de moyennes nulles et de second moment unité. Plus généralement, on peut utiliser des variables aléatoires centrées indépendantes et identiquement distribuées dont le second moment est 1. En effet, le processus aléatoire $Y_N(\tau) = X_{\lfloor \tau N \rfloor} / \sqrt{N}$ pour $t \in [0, 1]$ est constitué des sauts $s_i = \eta_i / \sqrt{N}$ de hauteur typique $\sqrt{\langle s_i^2 \rangle} = N^{-1/2}$ tous les $1/N$ pas de temps. Dans la limite d'un nombre de pas N infini, la taille typique des sauts ainsi que l'intervalle de temps entre chaque saut tendent simultanément vers 0 avec $\langle Y_N(\tau)^2 \rangle = t$ restant fini. Ainsi, lorsque N tend vers l'infini, le processus $Y_N(t) = X_{\lfloor \tau N \rfloor} / \sqrt{N}$ tend vers le brownien libre sur $\tau \in [0, 1]$, il s'agit du théorème de Donsker. On a tracé dans le volet **(a)** de la figure 2.1 un exemple de trajectoire $Y_N(t)$ obtenue en simulant une marche aléatoire constituée de $N = 100$ pas.

2.2 Les browniens contraints

Je considérerai par la suite des prolongements immédiats du mouvement brownien précédent :

- Le pont (figure 2.1 (b)) est un mouvement brownien conditionné à revenir à son point de départ $x_B(t) = x_B(0) = 0$. Avec $x(\tau)$ un brownien libre sur $[0, t]$, on peut construire le processus $x(\tau) - x(t)\tau/t$ qui est alors un pont sur $[0, t]$. Il est ainsi possible de simuler un pont à partir de la marche aléatoire X_i définie par (2.2) en posant

$$Y_i = X_i - \frac{i}{N}X_N, \quad i \in [0, N]. \quad (2.3)$$

Dans la limite où N tend vers l'infini, $Y_{\lfloor \tau N \rfloor} / \sqrt{N}$ converge vers un pont sur $[0, 1]$.

- L'excursion (figure 2.1 (c)) est un pont positif tout du long de l'intervalle considéré $x(\tau) \geq 0$ pour $\tau \in [0, t]$ avec $x(0) = x(t) = 0$.

Le processus $\sqrt{x_{B,1}(\tau)^2 + x_{B,2}(\tau)^2 + x_{B,3}(\tau)^2}$ construit à partir de trois ponts sur $[0, t]$ indépendants $\{x_{B,i}(\tau)\}_{i=1,2,3}$ est une excursion. En utilisant cette relation, il est possible de simuler une excursion

$$E_i = \sqrt{Y_{1,i}^2 + Y_{2,i}^2 + Y_{3,i}^2}, \quad i \in [1, N], \quad (2.4)$$

où les marches Y_1, Y_2 et Y_3 sont trois réalisations indépendantes de la marche aléatoire (2.3). $E_{\lfloor \tau N \rfloor} / \sqrt{N}$ converge, dans la limite où N tend vers l'infini, vers l'excursion sur $[0, 1]$.

- Le méandre (figure 2.1 (d)) est un mouvement brownien positif tout au long de l'intervalle considéré $x(\tau) \geq 0$ pour $\tau \in [0, t]$.

Le processus $\sqrt{x_{B,1}(\tau)^2 + x_{B,2}(\tau)^2 + x_{B,3}(\tau)^2 + ft}$ est un méandre avec $\{x_{B,i}(\tau)\}$ pour $i \in \{1, 2, 3\}$ trois ponts indépendants sur $[0, t]$ et f une variable aléatoire de Rayleigh de loi $p(f) = fe^{-f^2/2}$. En posant

$$M_i = \sqrt{Y_{1,i}^2 + Y_{2,i}^2 + (Y_{3,i} + f\frac{n}{N})^2}, \quad i \in [0, N], \quad (2.5)$$

avec Y_1, Y_2 et Y_3 trois réalisations indépendantes de la marche aléatoire (2.3) et f une variable aléatoire distribuée suivant la loi de Rayleigh, la marche aléatoire $M_{\lfloor \tau N \rfloor} / \sqrt{N}$ converge vers un méandre dans la limite où N tend vers l'infini.

- Le brownien réfléchi (figure 2.1 (e)) est la valeur absolue d'un brownien libre tandis que le pont réfléchi (figure 2.1 (f)) est la valeur absolue d'un pont.

Pour distinguer les différents cas, nous n'utiliserons aucun indice pour le brownien libre, les indices B pour le pont, E pour l'excursion et Me pour le méandre. Nous utiliserons également l'exposant R pour les trajectoires réfléchies.

2.3 Statistique d'extrême du mouvement brownien

L'étude de la statistique du maximum $x_{\max} = \max\{x(\tau), \tau \in [0, t]\}$ (ou du minimum) des trajectoires browniennes est un exemple à la fois non trivial et exactement soluble de statistique d'extrême de variables aléatoires fortement corrélées. Ces statistiques pour les différents browniens contraints ont été énormément étudiées depuis les travaux de Lévy [101] et possède de multiples applications. L'étude de cette statistique peut se faire simplement en utilisant les propagateurs contraints que nous introduisons dans la section 3.1. Pour les différents cas, la statistique du maximum $p_{\max}(x, t) = \Pr.[x_{\max} \leq x]$ est

donnée par

(a) Brownien libre :

$$p_{\max}(x, t) = \operatorname{erf}\left(\frac{x}{\sqrt{2t}}\right), \quad (2.6)$$

(b) Pont :

$$p_{\max, B}(x, t) = 1 - e^{-\frac{2x^2}{t}}, \quad (2.7)$$

(c) Excursion :

$$p_{\max, E}(x, t) = \frac{\sqrt{2\pi^5 t^3}}{x^3} \sum_{n=1}^{\infty} n^2 e^{-\frac{\pi^2}{2x^2} n^2 t}, \quad (2.8)$$

(d) Méandre :

$$p_{\max, Me}(x, t) = \frac{\sqrt{2^3 \pi t}}{x} \sum_{n=0}^{\infty} e^{-\frac{\pi^2}{2x^2} (2n+1)^2 t}, \quad (2.9)$$

(e) Brownien réfléchi :

$$p_{\max}^R(x, t) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} e^{-\frac{\pi^2}{8x^2} (2n+1)^2 t}, \quad (2.10)$$

(f) Pont réfléchi :

$$p_{\max, B}^R(x, t) = \frac{\sqrt{2\pi t}}{x} \sum_{n=0}^{\infty} e^{-\frac{\pi^2}{8x^2} (2n+1)^2 t}. \quad (2.11)$$

Remarquons que pour chacun des cas, la loi du maximum s'écrit comme une fonction d'échelle $p_{\max}(x, t) = p_{\max}(x/\sqrt{t}, 1)$ conforme à la relation d'autosimilarité du brownien. Il est intéressant d'étudier d'autres quantités en rapport avec cette distribution comme par exemple la loi jointe du maximum et de son temps d'atteinte, la statistique du maximum avant le premier passage, c'est-à-dire un mouvement brownien démarrant en $x(0) = x_0$ et d'une durée non fixée t qui est son temps de premier passage en 0 variant d'une trajectoire à une autre, ou encore l'extension à des modèles de n browniens (indépendants ou en interaction) [137, 138, 144]. En imaginant la variable $\tau \in [0, t]$ comme une variable d'espace que nous noterons $y \in [0, L]$, le mouvement brownien permet de modéliser des interfaces élastiques [103]. La position de l'interface étant invariante par translation globale, il est nécessaire de définir une origine. Il existe pour cela plusieurs méthodes non équivalentes [136]. Une première façon de faire est d'imposer la condition $x(0) = 0$ ce qui revient à considérer des trajectoires de browniens libres comme nous l'avons défini précédemment. Dans ce cas, la loi du maximum de l'interface est donnée par la loi de Rayleigh (2.7). Une autre définition usuelle est de retrancher à l'interface sa moyenne spatiale

$$h(y) = x(y) - \frac{1}{L} \int_0^L dy' x(y'). \quad (2.12)$$

Dans ce cas, le maximum de l'interface est donné par

$$h_{\max} = \max_{y \in [0, L]} h(y) = x_{\max} - \frac{1}{L} \int_0^L dy' x(y') = \frac{1}{L} \int_0^L (x_{\max} - x(y')) dy'. \quad (2.13)$$

Ce maximum est distribué suivant une loi d'Airy [106] comme nous le verrons pas la suite dans la section 5.2.2. Une application intéressante de ces résultats en statistiques est le *test de Kolmogorov-Smirnov* [96] pour déterminer si un échantillon suit bien une loi donnée. Supposons avoir N réalisations indépendantes de variables x_1, x_2, \dots, x_N générées suivant

la loi $p(u) = \text{Pr.}[x \leq u]$ continue. On considère alors la fonction de répartition empirique

$$p_N(u) = \frac{1}{N} \sum_{i=1}^N \theta(u - x_i) \quad (2.14)$$

où les fonctions θ sont des fonctions d'Heaviside valant 1 si leur argument est positif et 0 sinon. Dans la limite thermodynamique des grands N , cette fonction de répartition tend vers $p(X)$ et le théorème central limite indique que $B(u) = \sqrt{N}(p_N(u) - p(u))$ est une variable aléatoire gaussienne dont la fonction de corrélation à deux points est, pour $u' > u$

$$\begin{aligned} \langle B(u)B(u') \rangle &= N \langle (p_N(u) - p(u))(p_N(u') - p(u')) \rangle \\ &= N (\langle p_N(u)p_N(u') \rangle - p(u)p(u')) \\ &= \frac{1}{N} \sum_{i,j} \langle \theta(u - x_i)\theta(u' - x_j) \rangle - Np(u)p(u') \\ &= \frac{1}{N} \sum_i \langle \theta(u - x_i) \rangle + \frac{1}{N} \sum_{i \neq j} \langle \theta(u - x_i) \rangle \langle \theta(u' - x_j) \rangle - Np(u)p(u') \\ &= p(u) + (N-1)p(u)p(u') - Np(u)p(u') = p(u) - p(u)p(u'). \end{aligned} \quad (2.15)$$

En reparamétrisant cette fonction de corrélation avec la variable $\tau = p(u)$ qui varie entre 0 et 1, on trouve $\langle B(\tau)B(\tau') \rangle = \tau(1 - \tau')$ pour $\tau < \tau'$ qui coïncide donc avec la fonction de corrélation d'un pont brownien. Comme une variable aléatoire gaussienne est complètement caractérisée par sa fonction à deux points, $B(\tau)$ est donc un pont brownien sur $\tau \in [0, 1]$. On en déduit donc que le maximum de $|B(\tau)|$, qui est aussi le maximum de $|B(u)|$, suit la loi (2.11) pour $t = 1$ indépendamment de la loi $p(u)$

$$\text{Pr.} \left[\max_u \left(\sqrt{N} |p_N(u) - p(u)| \right) \leq x \right] = \frac{\sqrt{2\pi}}{x} \sum_{n=0}^{\infty} e^{-\frac{\pi^2}{8x^2}(2n+1)^2}. \quad (2.16)$$

Ce résultat est utilisé pour tester statistiquement si les réalisations sont générées suivant la loi donnée.

2.4 La densité à partir du maximum

Introduit en équation (5), la densité à partir du maximum $\rho_{\max}(r, t)$ peut-être décrite comme la densité dont l'origine est la variable aléatoire $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$ avec $x(\tau)$ un brownien sur $[0, t]$ (pouvant être contraint). Plus précisément, $\rho_{\max}(r, t)dr$ est le temps que passe la trajectoire brownienne dans l'intervalle $[x_{\max} - r - dr, x_{\max} - r]$ comme le montre la figure 2.2. Elle peut s'écrire sous la forme d'une fonctionnelle du maximum

$$\rho_{\max}(r, t) = \int_0^t \delta(x_{\max} - x(\tau) - r) d\tau, \quad (2.17)$$

normalisée de manière à avoir $\int_0^\infty \rho_{\max}(r, t) dr = t$. Remarquons que, puisque le brownien est autosimilaire, la densité à partir du maximum possède la loi d'échelle

$$\rho_{\max}(r, t) = \sqrt{t} \rho_{\max} \left(\frac{r}{\sqrt{t}}, 1 \right). \quad (2.18)$$

On cherche à décrire cette densité à partir du maximum pour les différents moments browniens : quels sont ses moments $\langle \rho_{\max}^k(r, t) \rangle$, où $\langle \dots \rangle$ représente la moyenne sur les trajectoires possibles du brownien considéré, voir quelle est sa loi $P_t(\rho, r) d\rho = \text{Pr.}[\rho_{\max}(r, t) \in$

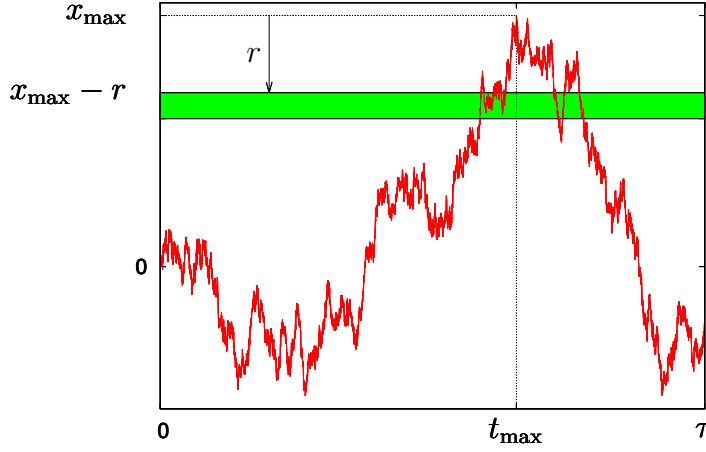


FIGURE 2.2: Exemple d’une réalisation d’un brownien libre $x(\tau)$ sur $[0, t]$. La trajectoire reste un temps $\rho_{\max}(r, t)dr$ (définie en (2.17)) à une distance comprise entre $[r, r + dr]$ de son maximum $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$ symbolisée par la bande verte.

$[\rho, \rho + d\rho]$. Même si les fonctionnelles du mouvement brownien ont déjà été largement étudiées [103], ce n’est pas le cas des fonctionnelles du maximum du mouvement brownien. En plus de la fonctionnelle du maximum (2.17), nous nous intéresserons également à toutes les fonctionnelles du maximum du type $\mathcal{O}_{\max}(t) = \int_0^t V(x_{\max} - x(\tau))d\tau$ avec $V(x)$ un potentiel quelconque, $\rho_{\max}(r, t)$ étant le cas important où $V(x) = \delta(x - r)$. Le premier moment $\langle \rho_{\max}(r, t) \rangle / t$ peut être interprété comme la densité de probabilité de trouver le mouvement brownien à la distance r du maximum. On peut alors exprimer la valeur moyenne des fonctionnelles du maximum comme

$$\langle \mathcal{O}_{\max}(t) \rangle = \langle \int_0^t V(x_{\max} - x(\tau))d\tau \rangle = \int_0^\infty \langle \rho_{\max}(r, t) \rangle V(r)dr, \quad (2.19)$$

ce qui assure un rôle clé à la densité à partir du maximum $\rho_{\max}(r, t)$. On étudiera par exemple dans le chapitre 5 le cas des potentiels $V(x) = x^\alpha$ qui dispose de plusieurs applications comme le calcul de l’aire sous un brownien [148] et le coût de l’algorithme optimal de recherche du maximum d’une marche aléatoire [119].

Chapitre 3

La méthode des propagateurs

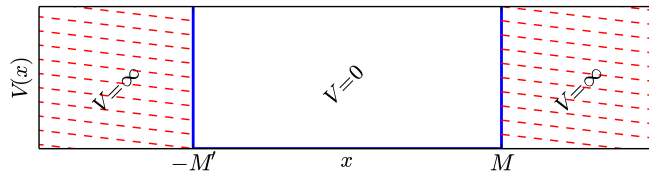
Dans ce chapitre, nous allons présenter une méthode simple fondée sur le comptage de chemin à l'aide des propagateurs que nous allons introduire en section 3.1 pour calculer la densité moyenne à partir du maximum $\langle \rho_{\max}(r, t) \rangle$ où $\langle X \rangle$ représente la moyenne de X sur les différentes trajectoires possibles. Nous illustrerons tout d'abord cette méthode sur le cas du brownien libre dans la section 3.2. Nous verrons ensuite dans la section 3.3 que cette méthode permet également de calculer les différents moments $\langle \rho_{\max, B}^k(r, t) \rangle$ dans le cas du pont et de l'excursion. Nous donnerons dans la section 3.4 les résultats dans le cas des autres browniens contraints que sont le méandre, le brownien libre réfléchi et le pont réfléchi. Nous étudierons alors les différentes distributions obtenues et comparerons à des simulations numériques. Nous généraliserons dans le chapitre suivant notre démarche à l'étude de la densité à partir du maximum après un temps fixé ainsi qu'au cas de plusieurs browniens (indépendants ou avec interaction). La plupart de ces résultats ont été publiés dans les articles [122, 123].

3.1 Propagateurs

Il sera nécessaire par la suite d'utiliser le concept de propagateur qui est la probabilité de trouver une trajectoire partant de α en β à $d\beta$ près après un temps t . Dans le cas du mouvement libre, ce propagateur est donné par la formule (2.1). Pour grouper les calculs de tous les browniens contraints, nous recherchons le propagateur d'une trajectoire confinée entre les bornes $[-M', M]$ avec des conditions de murs absorbants à chacun des bords et notons le propagateur recherché $G_{[-M', M]}(\alpha|\beta, t)$, les bornes M et M' pouvant être infinies. En utilisant le formalisme de Feynman-Kac, on peut reformuler cette étude en un problème de mécanique quantique où le hamiltonien est donné par

$$H = -\frac{1}{2} \frac{d^2}{dx^2} + V(x), \quad (3.1)$$

dont le potentiel $V(x)$ est un puits carré infini qui confine la particule quantique entre les bornes $[-M', M]$



$$V(x) = \begin{cases} +\infty & x < -M' \\ 0 & -M' < x < M \\ +\infty & M < x, \end{cases} \quad (3.2)$$

On peut facilement calculer les fonctions propres $\varphi_k(x)$ associées aux énergies propres E_k du hamiltonien (3.1)

$$\varphi_n(x) = \sqrt{\frac{2}{M+M'}} \sin\left(n\pi \frac{x+M'}{M+M'}\right), \quad E_n = \frac{1}{2} \left(\frac{n\pi}{M+M'}\right)^2 \quad n \in \mathbb{N}^*. \quad (3.3)$$

En utilisant la décomposition sur cette base d'état propre, le propagateur s'écrit

$$\begin{aligned} G_{[-M',M]}(\alpha|\beta, t) &= \langle \beta | e^{-Ht} | \alpha \rangle = \sum_{n=1}^{\infty} \varphi_n^*(\beta) \varphi_n(\alpha) e^{-E_n t} \\ &= \sum_{n=1}^{\infty} \frac{2}{M+M'} \sin\left(n\pi \frac{\alpha+M'}{M+M'}\right) \sin\left(n\pi \frac{\beta+M'}{M+M'}\right) e^{-\frac{\pi^2}{2(M+M')^2} n^2 t}. \end{aligned} \quad (3.4)$$

Par la suite, nous verrons que la transformée de Laplace du propagateur par rapport à t joue un rôle important. Celle ci s'écrit

$$\begin{aligned} \tilde{G}_{[-M',M]}(\alpha|\beta, s) &= \int_0^{\infty} dt G_{[-M',M]}(\alpha|\beta, t) e^{-st} \\ &= \sum_{n=1}^{\infty} \frac{2}{M'+M} \frac{\sin\left(n\pi \frac{\alpha+M'}{M+M'}\right) \sin\left(n\pi \frac{\beta+M'}{M+M'}\right)}{s + \frac{\pi^2}{2(M+M')^2} n^2}. \end{aligned} \quad (3.5)$$

En utilisant la formule valable pour $0 \leq x \leq 2\pi$

$$\sum_{k=1}^{\infty} \frac{\cos(kx)}{a^2 + k^2} = \frac{1}{2a} \left(\frac{\pi \cosh((\pi - x)a)}{\sinh(\pi a)} - \frac{1}{a} \right), \quad (3.6)$$

on peut resommer la transformée de Laplace et on trouve finalement

$$\tilde{G}_{[-M,M']}(\alpha|\beta, s) = \frac{2 \sinh\left(\sqrt{2s}(M - \max[\alpha, \beta])\right) \sinh\left(\sqrt{2s}(M' + \min[\alpha, \beta])\right)}{\sqrt{2s} \sinh\left(\sqrt{2s}(M + M')\right)}. \quad (3.7)$$

Lorsque l'on fait tendre la borne M' vers l'infini, on obtient le propagateur du brownien seulement contraint à rester sous la borne M durant tout l'intervalle $[0, t]$. Ce propagateur, que nous notons simplement $G_M(\alpha|\beta, t)$ pour alléger la notation, peut également être calculé en utilisant la méthode des images à partir du propagateur pour le cas libre (2.1)

$$G_M(\alpha|\beta, t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{(\beta-\alpha)^2}{2t}} - e^{-\frac{(2M-\beta-\alpha)^2}{2t}} \right). \quad (3.8)$$

dont la transformée de Laplace est

$$\tilde{G}_M(\alpha|\beta, s) = \frac{1}{\sqrt{2s}} \left(e^{-\sqrt{2s}|\beta-\alpha|} - e^{-\sqrt{2s}(2M-\beta-\alpha)} \right). \quad (3.9)$$

Nous dénoterons aussi pour alléger la notation $G_+(\alpha|\beta, t)$ le cas où la trajectoire est confinée sur l'axe positif c'est-à-dire $[-M', M] = [0, \infty[$. Dans ce cas, le propagateur et sa transformée de Laplace se réécrivent

$$G_+(\alpha|\beta, t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{(\alpha-\beta)^2}{2t}} - e^{-\frac{(\alpha+\beta)^2}{2t}} \right), \quad (3.10)$$

$$\tilde{G}_+(\alpha|\beta, s) = \frac{1}{\sqrt{2s}} \left(e^{-\sqrt{2s}|\alpha-\beta|} - e^{-\sqrt{2s}(\alpha+\beta)} \right), \quad (3.11)$$

résultats que l'on peut facilement retrouver en utilisant la méthode des images à partir du propagateur pour le cas libre (2.1).

3.2 Illustration de la méthode : calcul de $\langle \rho_{\max}(r, t) \rangle$ dans le cas du brownien libre

Nous allons tout d'abord illustrer la méthode sur l'exemple du brownien libre. Rappelons que $\rho_{\max}(r, t)dr$ correspond au temps cumulé que le brownien passe à une distance entre r et $r + dr$ de son maximum. Raisonnons tout d'abord sur une trajectoire de maximum M sur $[0, t]$ atteint au temps t_{\max} et de position finale $x(t) = x_F$. Nous terminerons notre raisonnement en moyennant sur ces 3 paramètres c'est-à-dire en intégrant sur toutes les valeurs possibles de ces paramètres à la fin de notre calcul. Pour cette trajectoire, $\rho_{\max}(r, t)$ compte le "nombre de fois" que passe le brownien à la position $M - r$. On cherche donc à compter de telles trajectoires qui passent en $M - r$ au temps τ en intégrant sur toutes les possibilités de temps $\tau : 0 \leq \tau \leq t$. En effet, si une trajectoire passe plusieurs fois en $M - r$, elle y passe en des temps τ différents et est donc comptée le bon nombre de fois grâce à l'intégration sur toutes les valeurs de τ possibles. Il ne reste alors qu'à intégrer sur toutes les valeurs permises de x_F , M et t_{\max} . Néanmoins, on ne peut pas imposer au mouvement brownien simultanément $x(t_{\max}) = M$ et $x(t) < M$ juste avant et après t_{\max} . Pour éviter cette difficulté [105], on régularise en imposant $x(t_{\max}) = M - \varepsilon$ et l'on prendra la limite $\varepsilon \rightarrow 0$ à la fin du calcul. Il ne reste alors qu'à intégrer sur les paramètres libres M , x_F and t_{\max} et renormaliser par la constante $Z(\varepsilon)$ qui compte le nombre de chemins régularisés. Dans chacun des trois intervalles délimités par τ , t_{\max} et t , le brownien est contraint à rester sous la borne M . Ce nombre de chemins est donné par le propagateur $G_M(\alpha|\beta, t)$ (3.8) qui donne le nombre de chemins démarrant en $x = \alpha < M$ terminant en $x = \beta < M$ après un temps t tout en restant en dessous de la borne M durant tout l'intervalle $[0, t]$. Comme le montre la figure 3.1, lorsque l'on divise l'intervalle temporel $[0, t]$ en trois parties, il faut considérer deux cas : $\tau < t_{\max}$ et $\tau > t_{\max}$ qui contribue à $\langle \rho_{\max}(r, t) \rangle$. Chacune des contributions fait intervenir un produit de 3 propagateurs (notés sur la figure 3.1) en utilisant la propriété de Markov du mouvement brownien. Il ne reste alors qu'à intégrer chacune des contributions sur les paramètres libres $x_F < M$, $M > 0$, τ et t_{\max} . Ces deux configurations apportent en réalité la même contribution à $\langle \rho_{\max}(r, t) \rangle$: en effet, si l'on translate la trajectoire $x(\tau) \rightarrow x(\tau) - x_F/2$ (pour chaque réalisation du brownien), les deux contributions ci-dessus sont alors symétriques l'une de l'autre par renversement du temps. Ainsi, on obtient finalement :

$$\begin{aligned} \langle \rho_{\max}(r, t) \rangle &= \lim_{\varepsilon \rightarrow 0} \frac{2}{Z(\varepsilon)} \int_0^\infty dM \int_0^t dt_{\max} \int_{-\infty}^M dx_F \int_0^{t_{\max}} d\tau G_M(0|M - r, \tau) \\ &\quad \times G_M(M - r|M - \varepsilon, t_{\max} - \tau) G_M(M - \varepsilon|x_F, t - t_{\max}), \end{aligned} \quad (3.12)$$

Le facteur 2 provenant des deux configurations équivalentes $\tau < t_{\max}$ et $\tau > t_{\max}$. Dans l'équation (3.12), la constante de normalisation $Z(\varepsilon)$ comptant le nombre de trajectoires régularisées est donnée par

$$Z(\varepsilon) = \int_0^\infty dM \int_0^t dt_{\max} \int_{-\infty}^M dx_F G_M(0|M - \varepsilon, t_{\max}) G_M(M - \varepsilon|x_F, t - t_{\max}). \quad (3.13)$$

Celle-ci peut être facilement calculée à partir du propagateur (3.8) dans la limite $\varepsilon \rightarrow 0$

$$Z(\varepsilon) \underset{\varepsilon \rightarrow 0}{\sim} 2\varepsilon^2. \quad (3.14)$$

Dans (3.12), on reconnaît une structure de convolution, ce qui nous pousse à prendre la transformée de Laplace de $\langle \rho_{\max}(r, t) \rangle$ par rapport au temps t

$$\begin{aligned} \int_0^\infty dt e^{-st} \langle \rho_{\max}(r, t) \rangle &= \lim_{\varepsilon \rightarrow 0} \frac{2}{Z(\varepsilon)} \int_0^\infty dM \int_{-\infty}^M dx_F \tilde{G}_M(0|M - r, s) \\ &\quad \tilde{G}_M(M - r|M - \varepsilon, s) \tilde{G}_M(M - \varepsilon|x_F, s). \end{aligned} \quad (3.15)$$

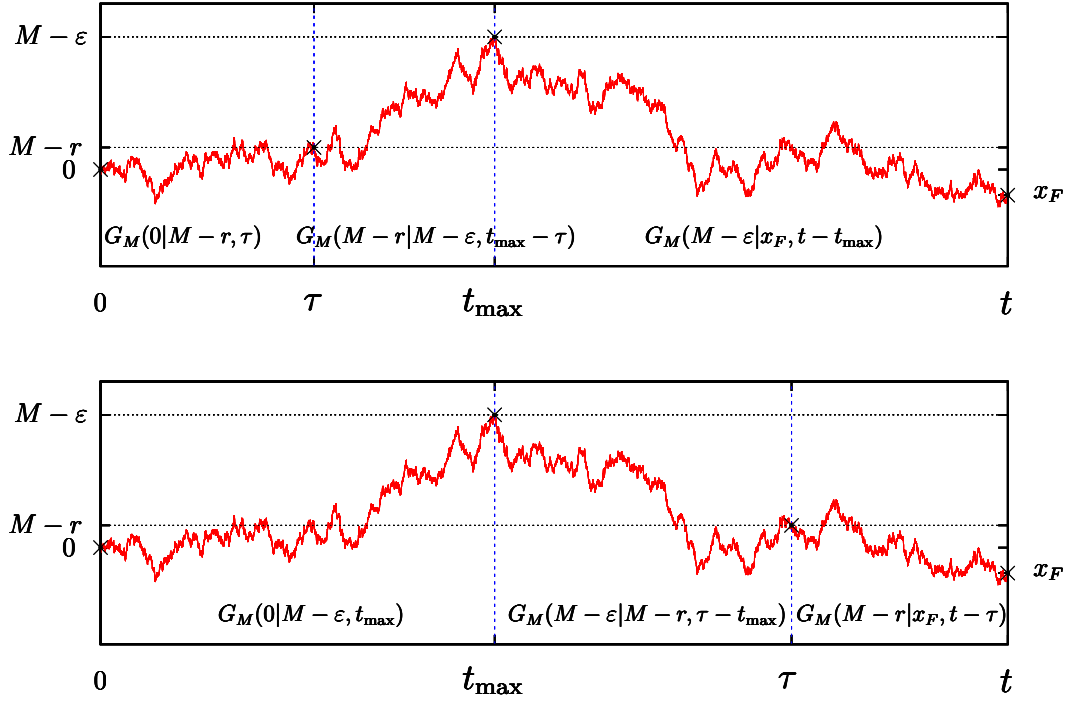


FIGURE 3.1: Les deux contributions à $\langle \rho_{\max}(r, t) \rangle$ dans le cas du brownien libre. Le mouvement brownien démarre en 0, passe en $M - r$ au temps τ ainsi qu'en $M - \varepsilon$ au temps t_{\max} et finit en x_F au temps t . Il suffit alors d'intégrer chacune des contributions sur les paramètres libres $x_F < M$, $M > 0$, τ et t_{\max} .

où $\tilde{G}_M(\alpha|\beta, s)$ est la transformée de Laplace par rapport à t du propagateur (3.8) que nous avons calculé en (3.9). Cette relation se comprend aisément à partir de la première trajectoire de la figure 3.1, la seconde donnant la même contribution. Il est ainsi possible à partir de n'importe quelle trajectoire de facilement calculer sa contribution à la densité moyenne à partir du maximum dans le domaine de Laplace. Ce calcul à l'aide des propagateurs devient alors un simple jeu où l'on multiplie les propagateurs dans le domaine de Laplace pour obtenir l'analogue de la relation (3.15) pour la trajectoire considérée. Pour terminer le calcul, on utilise la transformée de Laplace du propagateur (3.9), dans la limite ε vers 0, on obtient directement

$$\begin{aligned} \int_0^\infty dt e^{-st} \langle \rho_{\max}(r, t) \rangle &= 8 \left[\int_0^r dM \frac{e^{-\sqrt{2s}r}}{\sqrt{2s}} \sinh(\sqrt{2s}M) e^{-\sqrt{2s}r} \int_{-\infty}^M dx_F e^{-\sqrt{2s}(M-x_F)} \right. \\ &\quad \left. + \int_r^\infty dM \frac{e^{-\sqrt{2s}M}}{\sqrt{2s}} \sinh(\sqrt{2s}r) e^{-\sqrt{2s}r} \int_{-\infty}^M dx_F e^{-\sqrt{2s}(M-x_F)} \right]. \end{aligned} \quad (3.16)$$

Les deux dernières intégrales sur x_F et M peuvent être calculées et on trouve

$$\int_0^\infty dt e^{-st} \langle \rho_{\max}(r, t) \rangle = 8 \frac{e^{-\sqrt{2s}r} - e^{-2\sqrt{2s}r}}{(2s)^{3/2}}. \quad (3.17)$$

Après une transformée de Laplace inverse, on obtient le résultat

— Densité moyenne à partir du maximum du brownien libre —

$$\langle \rho_{\max}(r, t) \rangle = 8\sqrt{t} \left(\Phi^{(2)}\left(\frac{r}{\sqrt{t}}\right) - \Phi^{(2)}\left(2\frac{r}{\sqrt{t}}\right) \right), \quad (3.18)$$

où $\Phi^{(2)}(x)$ appartient à la famille de fonctions $\Phi^{(j)}(x)$ (A.5) introduite dans l'appendice A.

Cette méthode peut en principe être adaptée pour calculer les différents moments de la densité $\langle \rho_{\max}^k(r, t) \rangle$ en comptant les chemins passant k fois en $M - r$. Ce calcul est néanmoins lourd dès que $k > 1$. Dans le cas du pont et de l'excursion, comme nous allons le voir dans la partie suivante, une identité permet de calculer les différents moments plus simplement. Nous verrons dans le chapitre suivant une approche plus rapide donnant directement la fonction génératrice de $\rho_{\max}(r, t)$. La méthode des propagateurs permet néanmoins de calculer simplement la densité moyenne à partir du maximum pour les différents browniens contraints.

3.3 Calcul des différents moments $\langle \rho_{\max, B}^k(r, t) \rangle$ pour le pont et l'excursion

Nous allons dans cette partie calculer les différents moments de la densité à partir du maximum pour le pont et l'excursion. Dans ce cas, un argument permet de montrer que les deux densités sont en réalité égales et sont toutes deux le temps local d'une excursion c'est-à-dire la densité. Ceci permet de simplifier drastiquement le calcul. Introduisons tout d'abord la transformation de Vervaat [161] : prenons une trajectoire de pont $x_B(\tau)$ sur $\tau \in [0, t]$, c'est-à-dire un mouvement brownien qui démarre et termine à la position 0 comme sur la figure 3.2 (a). Permutons alors les parties antérieures et postérieures à un temps T pour créer une nouvelle trajectoire $y_{x_B, T}(\tau)$, la continuité de ce nouveau chemin est garantie grâce aux conditions aux limites $x(t) = x(0) = 0$ imposées au pont

$$y_{x_B, T}(\tau) = \begin{cases} x_B(\tau + T) - x_B(T) & \text{pour } \tau < t - T, \\ x_B(T - t + \tau) - x_B(T) & \text{pour } \tau > t - T. \end{cases} \quad (3.19)$$

Si on choisit pour T le temps t_{\max} auquel le pont $x_B(\tau)$ atteint son maximum, la trajectoire $-y_{x, t_{\max}}(\tau)$ que l'on obtient est une excursion comme le montre la figure 3.2 (b). La transformation inverse est obtenue en redécoupant une excursion à un temps aléatoire T tiré uniformément sur $[0, t]$. Dans cette transformation, la densité à partir du maximum $\rho_{\max, B}(r, t)$, qui correspond au temps cumulé que le brownien passe à une distance entre r et $r + dr$ de son maximum, symbolisé par la bande verte sur la figure, est devenu le temps local de l'excursion, temps passé en r . De plus, si l'on permute à partir de $T = t_{\min}$ comme le montre la figure 3.2 (c), avec t_{\min} le temps auquel le brownien atteint son minimum, on obtient encore une fois une excursion mais où $\rho_{\max, B}(r, t)$ a été transformé en la densité à partir du maximum de l'excursion $\rho_{\max, E}(r, t)$. Ceci montre que ces trois quantités sont égales en loi

$$\rho_{\max, B}(r, t) \stackrel{\text{loi}}{=} \rho_{\max, E}(r, t) \stackrel{\text{loi}}{=} T_{\text{loc}, E}(r, t) = \int_0^t \delta(x_E(\tau) - r) d\tau, \quad (3.20)$$

où $x_E(\tau)$ est une excursion brownienne. Cette identité permet de simplifier grandement les calculs des différents moments $\langle \rho_{\max, B}^k(r, t) \rangle$ de la densité à partir du maximum dans les cas du pont et de l'excursion puisque le temps local est une fonctionnelle du mouvement brownien qui ne fait plus apparaître la notion de maximum.

$$\langle \rho_{\max, B}^k(r, t) \rangle = \langle \rho_{\max, E}^k(r, t) \rangle = \langle T_{\text{loc}, E}^k(r, t) \rangle = \left\langle \prod_{i=1}^k \left(\int_0^t dt_i \delta(x_E(t_i) - r) \right) \right\rangle. \quad (3.21)$$

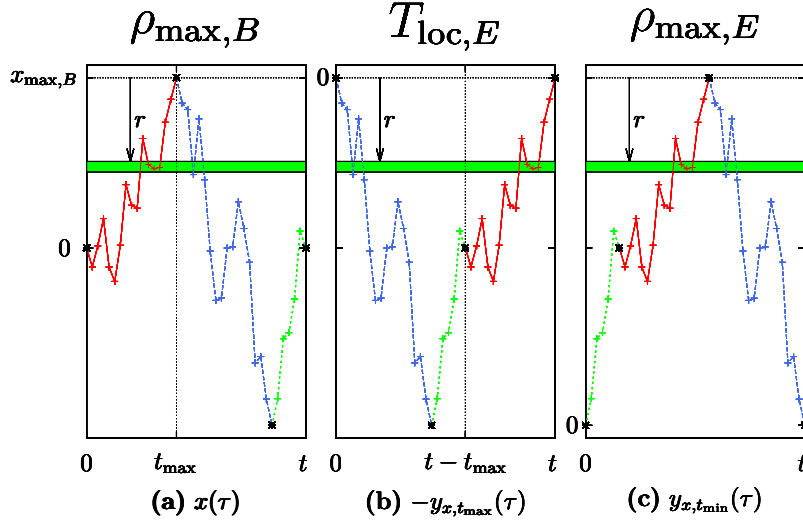


FIGURE 3.2: Transformation de Vervaat qui montre que la densité à partir du maximum pour le pont est identique en loi à celle de l'excursion et également au temps local de l'excursion. **(a)** Une trajectoire du pont brownien $x(\tau)$. Le temps cumulé dans la bande verte entre $[x_{\max} - r - dr, x_{\max} - r]$ est $\rho_{\max,B}(r, t)dr$. **(b)** La trajectoire $-y_{x,t_{\max}}(\tau)$ obtenue après la transformation de Vervaat en découpant au maximum est une excursion. Le temps cumulé dans la bande verte est désormais $T_{\text{loc},E}(r, t)dr$. **(c)** La trajectoire $y_{x,t_{\min}}(\tau)$ obtenue après la transformation de Vervaat en découpant au minimum est également une excursion. Le temps cumulé dans la bande verte est désormais $\rho_{\max,E}(r, t)dr$.

Comme nous l'avons fait précédemment dans le cas du brownien libre en équation (3.12), les expressions des différents moments (3.21) peuvent être calculées en utilisant les propagateurs $G_+(\alpha|\beta, t)$ donné en (3.10) du mouvement brownien qui démarre en $\alpha > 0$ au temps 0 pour terminer en $\beta > 0$ au temps t tout en restant positif sur tout l'intervalle $[0, t]$, condition imposée par l'excursion. Il est nécessaire encore une fois [105] d'introduire un régulateur $x_E(0) = x_E(t) = \varepsilon > 0$ puis de prendre la limite $\varepsilon \rightarrow 0$ à la fin de notre calcul, comme nous l'avons déjà fait dans l'équation (3.12) car il n'est pas possible d'imposer à la fois les conditions $x_E(0) = 0$ et $x_E(0^+) > 0$. Nous avons alors

$$\begin{aligned} \langle \rho_{\max,B}^k(r, t) \rangle &= \langle \rho_{\max,E}^k(r, t) \rangle = \langle T_{\text{loc},E}^k(r, t) \rangle \\ &= \lim_{\varepsilon \rightarrow 0} \frac{k!}{Z_B(\varepsilon)} \int_{0 < t_1 < t_2 < \dots < t_k < t} dt_1 \dots dt_k G_+(\varepsilon|r, t_1) G_+(r|r, t_2 - t_1) \dots G_+(r|\varepsilon, t - t_k), \end{aligned} \quad (3.22)$$

où le facteur combinatoire $k!$ provient des différentes permutations des temps intermédiaires t_1, \dots, t_k où l'excursion brownienne $x_E(\tau)$ atteint la valeur r . La constante de normalisation qui compte les chemins régularisés $Z_B(\varepsilon)$ est donnée par

$$Z_B(\varepsilon) = G_+(\varepsilon|\varepsilon, t) \underset{\varepsilon \rightarrow 0}{\sim} \varepsilon^2 \sqrt{\frac{2}{\pi}} t^{-3/2}. \quad (3.23)$$

Comme dans le cas précédent, on reconnaît dans l'intégrale multiple sur les temps t_1, \dots, t_k de l'équation (3.22) un produit de convolution. Il est ainsi intéressant de prendre la transformée de Laplace de l'intégrale multiple par rapport au temps t

$$\begin{aligned} &\int_0^\infty dt e^{-st} \int_{0 < t_1 < t_2 < \dots < t_k < t} dt_1 \dots dt_k G_+(\varepsilon|r, t_1) G_+(r|r, t_2 - t_1) \dots G_+(r|\varepsilon, t - t_k) \\ &= \tilde{G}_+(\varepsilon|r, s)^2 \tilde{G}_+(r|r, s)^{k-1} \underset{\varepsilon \rightarrow 0}{\sim} 4\varepsilon^2 e^{-2\sqrt{2}sr} \frac{(1 - e^{-2\sqrt{2}sr})^{k-1}}{(2s)^{(k-1)/2}}, \end{aligned} \quad (3.24)$$

où l'on a utilisé l'expression de la transformée de Laplace du propagateur (3.11). Après une inversion de Laplace, on obtient en utilisant la constante de normalisation (3.23) les différents moments de la densité à partir du maximum

— Moments de la densité à partir du maximum pour le pont —

$$\begin{aligned}
 \langle \rho_{\max, B}^k(r, t=1) \rangle &= \langle \rho_{\max, E}^k(r, t=1) \rangle = \langle T_{\text{loc}, E}^k(r, t=1) \rangle \\
 &= 2\sqrt{2\pi} k! \sum_{l=1}^k (-1)^{l+1} \binom{k-1}{l-1} \Phi^{(k-2)}(2rl), \quad (3.25)
 \end{aligned}$$

avec la convention $\Phi^{(-1)} = -\partial_r \Phi^{(0)}$ et où les $\Phi^{(j)}$ sont définies dans l'appendice A. La dépendance temporelle $\langle \rho_{\max, B}^k(r, t) \rangle = t^{n/2} \langle \rho_{\max, B}^k(\frac{r}{\sqrt{t}}, 1) \rangle$ découle directement de l'autosimilarité du brownien. Nous réobtenons ainsi le résultat de Takács [150] en équation (38). Pour $k=1$, on trouve ainsi que le densité à partir du maximum pour un pont sur l'intervalle $[0, 1]$ est [150]

$$\langle \rho_{\max, B}(r, 1) \rangle = 4re^{-2r^2}. \quad (3.26)$$

Remarquons que dans ce cas, cela coïncide avec la densité de probabilité du maximum du pont (2.7). Ce résultat est une propriété générale pour un signal périodique $x(t) = x(0)$ [24], en réutilisant le découpage introduit en (3.19) avec un T fixé, la moyenne sur tous les ponts $x_B(\tau)$ possibles permet aux trajectoires $y_{x_B, T}(\tau)$ d'explorer également l'ensemble des trajectoires des ponts possibles. On trouve alors

$$\begin{aligned}
 \langle \rho_{\max, B}(r, t) \rangle &= \int_0^t d\tau \langle \delta(x_{\max, B} - x_B(\tau) - r) \rangle \\
 &= \langle \delta(\max_{\tau \in [0, t]} y_{x_B, T}(\tau) - r) \rangle \\
 &= p'_{\max, B}(r, t). \quad (3.27)
 \end{aligned}$$

3.4 Résultats pour les autres browniens contraints

En utilisant la méthode (3.12), il est également possible de calculer la densité moyenne à partir du maximum pour les autres browniens contraints comme le méandre (cf. Fig. 2.1 (d)), le brownien libre réfléchi (cf. Fig. 2.1 (e)) et le pont réfléchi (cf. Fig. 2.1 (f)). Il suffit dans ce cas de décomposer comme précédemment les différentes trajectoires pour toutes les configurations possibles en divisant l'intervalle de temps $[0, t]$ et en étudiant pour chacune des parties temporelles dans quelles régions de l'espace est confiné le brownien. La propriété de Markov permet alors d'écrire chaque contribution comme un produit de convolution qui se simplifie grandement dans le domaine de Laplace. Nous pouvons alors faire le même travail que précédemment en utilisant le propagateur dans le domaine de Laplace (3.7). On obtient, après des calculs que l'on peut trouver dans [123], les différentes densités moyennes sur l'intervalle $[0, 1]$

— Densité moyenne à partir du maximum du méandre —

$$\langle \rho_{\max, Me}(r, 1) \rangle = 2\sqrt{2\pi} \left(\sum_{n=1}^{\infty} \frac{4n(-1)^n}{2n^2 + 3(-1)^n - 5} \Phi^{(1)}(nr) - \Phi^{(1)}(2r) \right), \quad (3.28)$$

	$r \sim 0$	$r \rightarrow \infty$	$\langle r \rangle$	r_{typ}	simulations	
					$\langle r \rangle$	r_{typ}
Brownien libre	$4r$	$\sqrt{\frac{2}{\pi}} \frac{4}{r^2} e^{-\frac{r^2}{2}}$	$\sqrt{\frac{2}{\pi}} = 0.7979\dots$	0.5145...	0.79	0.52
Pont et Excursion	$4r$	$4r e^{-2r^2}$	$\frac{1}{2} \sqrt{\frac{\pi}{2}} = 0.6267\dots$	0.5	0.62	0.48
Méandre	$4r$	$\frac{4}{3r} e^{-\frac{r^2}{2}}$	$\sqrt{\frac{\pi}{2}} (2 \log 2 - \frac{3}{4}) = 0.7975\dots$	0.4907...	0.79	0.48
Brownien libre réfléchi	$4r$	$\sqrt{\frac{2}{\pi}} \frac{8}{3r^2} e^{-\frac{r^2}{2}}$	$\frac{3\pi-4}{3\sqrt{2}\pi} = 0.7214\dots$	0.5212...	0.72	0.52
Pont réfléchi	$4r$	$\frac{3}{r} e^{-2r^2}$	$\sqrt{\frac{\pi}{2}} (\log 2 - \frac{1}{4}) = 0.5554\dots$	0.4907...	0.55	0.48

TABLE 3.1: Caractéristiques de la densité moyenne à partir du maximum pour les différents mouvements browniens contraints et comparaison aux simulations. Les valeurs simulées obtenues après avoir moyenné sur 10^4 trajectoires de 10^4 pas sont données avec une barre d'erreur statistique de ± 0.01 pour $\langle r \rangle$ et ± 0.02 pour r_{typ} .

— Densité moyenne à partir du maximum du brownien libre réfléchi —

$$\langle \rho_{\text{max}}^R(r, 1) \rangle = 8 \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{-3 - 2n + 12n^2 + 8n^3} \left(4n^2(3 + 2n)\Phi^{(2)}(2nr) + (2n + 1)^2(3 - (2n + 1)^2)\Phi^{(2)}((2n + 1)r) \right), \quad (3.29)$$

— Densité moyenne à partir du maximum du pont réfléchi —

$$\langle \rho_{\text{max},B}^R(r, 1) \rangle = 2\sqrt{2\pi} \left(4 \sum_{n=0}^{\infty} n(-1)^{n+1}\Phi^{(1)}(2nr) - \Phi^{(1)}(2r) \right), \quad (3.30)$$

où $\Phi^{(1)}(x)$ et $\Phi^{(2)}(x)$ appartiennent à la famille de fonctions définie dans l'appendice A, la dépendance temporelle étant donnée par la relation 2.18. Remarquons que pour chacun des cas, la densité moyenne à partir du maximum a un comportement différent du cas des variables indépendantes et identiquement distribuées [139] que nous avons rappelées en introduction. Dans tous les cas, $\langle \rho_{\text{max}}(r, t) \rangle \underset{r \rightarrow 0}{\sim} 4r$ car le voisinage du maximum n'est pas sensible aux conditions aux bords. Le comportement à grand r dépend quant à lui des conditions aux bords, cf. colonne 3 du tableau 3.1. Remarquons qu'il se comporte comme $\propto e^{-r^2/2}$ pour les browniens qui ont leur extrémité libre et comme $\propto e^{-2r^2}$ pour les browniens qui ont la condition $x(t) = 0$. On compare en figure 3.3 les différentes formules - les différentes lignes sur la figure - aux données obtenues en simulant 10^4 marches aléatoires de taille 10^4 - les différents points sur la figure - (cf. introduction 2.2). Dans tous les cas, il y a un très bon accord entre nos formules et les données numériques. Nous donnons en colonne 4 du tableau 3.1 la distance moyenne du brownien à son maximum :

$$\langle r \rangle = \int_0^{\infty} dr \, r \langle \rho_{\text{max}}(r, 1) \rangle. \quad (3.31)$$

Nous donnons également en colonne 5 du tableau 3.1 la valeur typique r_{typ} de la distance du brownien à son maximum c'est-à-dire la valeur de r qui rend maximal $\langle \rho_{\text{max}}(r, 1) \rangle$. Les colonnes 6 et 7 du tableau 3.1 donnent les résultats pour ces quantités obtenues après une moyenne sur 10^4 réalisations de marches aléatoires de 10^4 pas et sont en total accord avec nos résultats.

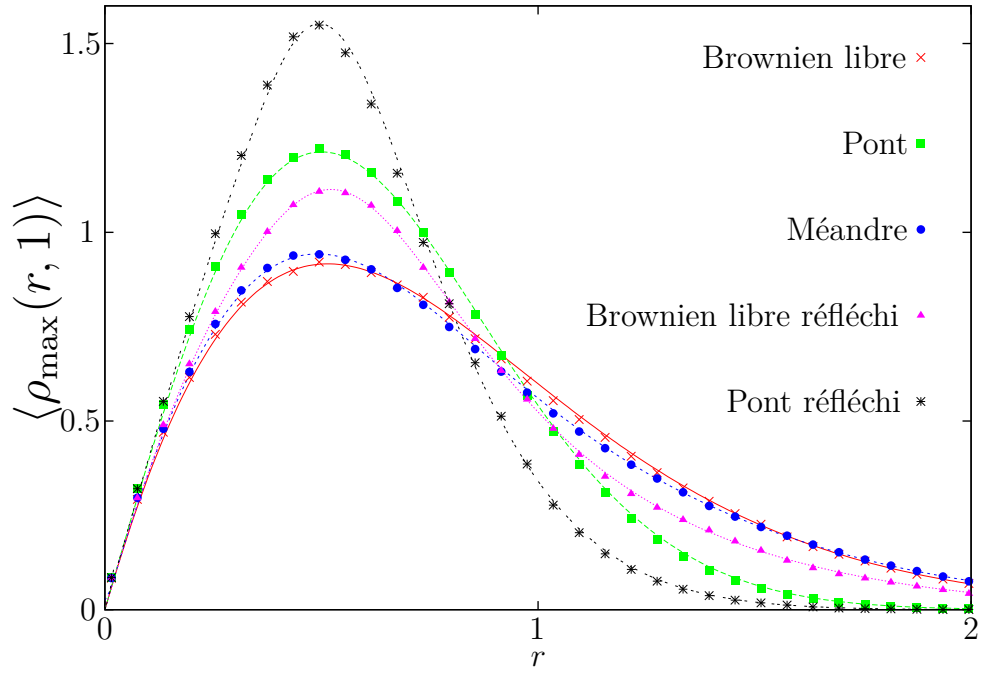


FIGURE 3.3: Résultats numériques obtenus en moyennant sur 10^4 marches aléatoires de 10^4 pas pour les différents mouvements browniens contraints (points). Les algorithmes de génération des browniens sont détaillés dans l'introduction 2.2. Les lignes représentent nos résultats analytiques donnés par les différentes équations (3.18), (3.26), (3.28), (3.29) et (3.30) et correspondent avec nos simulations.

Chapitre 4

Extensions : étude temporelle et cas de plusieurs browniens

Dans ce chapitre, nous généralisons dans la section 4.1 notre calcul de la densité à partir du maximum en considérant la densité moyenne à une distance r du maximum et à un temps θ du temps où le brownien atteint son maximum pour les cas du brownien libre et du pont. Nous nous intéressons ensuite à la densité à partir du maximum pour plusieurs ponts indépendants dans la section 4.2.1 et également pour plusieurs ponts contraints à ne pas se croiser dans la section 4.2.2. Certaines étapes des calculs sont données en appendice B et C.

4.1 Étude temporelle

Dans le cas du mouvement brownien, il est possible d'étendre notre étude de la densité à partir des extrêmes en considérant l'aspect temporel. Cette piste de recherche est la suite directe de l'étude de la loi jointe du maximum et de son temps d'attente [138]. En plus de rechercher le temps que passe le brownien à une distance r de son maximum, il est naturel de se poser des questions telles que "À quel moment le brownien a-t-il passé ce temps à r du maximum? Est-ce pour des instants proches ou lointains du maximum?". Nous introduisons ainsi, sur le modèle de la définition (2.17), la densité à partir du maximum à un temps fixé pour un brownien sur $[0, t]$ qui, comme le montre la figure 4.1, n'explore le brownien qu'au temps $t_{\max} + \theta$ avec t_{\max} le temps auquel le brownien atteint son maximum $x_{\max} = \max\{x(\tau), \tau \in [0, t]\}$. Nous ne considérerons dans ce cas que le premier moment qui, similairement au raisonnement (2.19), peut être interprété comme la densité de probabilité d'être à une distance r et à une différence de temps $\theta \in [-t, t]$ du maximum. Nous noterons par la suite cette densité de probabilité

$$\Xi(r, \theta, t) = \langle \delta(x_{\max} - x(t_{\max} + \theta) - r) \rangle. \quad (4.1)$$

Cette quantité peut être déterminé en utilisant le produit de propagateur comme nous l'avons fait pour l'équation (3.12). En intégrant sur la variable θ , avec $\theta \in [-t, t]$, on retrouve la densité à partir du maximum

$$\int_{-t}^t d\theta \Xi(r, \theta, t) = \langle \int_{-t}^t d\theta \delta(x_{\max} - x(t_{\max} + \theta) - r) \rangle = \langle \rho_{DOS}(r, t) \rangle, \quad (4.2)$$

ce qui nous donne la condition de normalisation

$$\int_0^\infty dr \int_{-t}^t d\theta \Xi(r, \theta, t) = t. \quad (4.3)$$

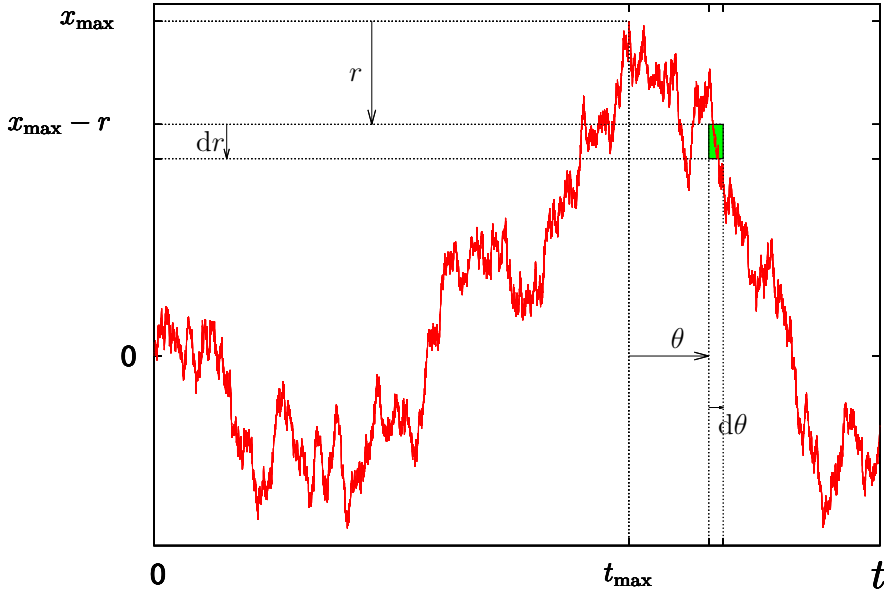


FIGURE 4.1: La densité à partir du maximum à un temps θ fixé calcule la probabilité que la trajectoire soit à une distance dans l'intervalle $[r, r + dr]$ de son maximum $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$ après un temps θ , c'est-à-dire si la trajectoire passe dans le carré vert.

4.1.1 Cas du brownien libre

Le cas du brownien libre s'obtient en modifiant l'équation (3.12). La symétrie par renversement du temps couplée à une translation de $x_F = x(t)$ discutée à propos de l'équation (3.12), comme le montre la figure 3.1, démontre que le cas $\theta < 0$ est identique au cas des $\theta > 0$. Dans le cas $\theta > 0$, on trouve

$$\begin{aligned} \Xi(r, \theta, t) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{Z(\varepsilon)} \int_0^\infty dM \int_0^{t-\theta} dt_{\max} \int_{-\infty}^M dx_F G_M(0|M-\varepsilon, t_{\max}) \\ &\quad G_M(M-\varepsilon|M-r, \theta) G_M(M-r|x_F, t-t_{\max}-\theta), \end{aligned} \quad (4.4)$$

On retrouve une structure en produit de convolution. On utilise alors la transformée de Laplace de l'intégrale par rapport à la variable $T = t - \theta$

$$\begin{aligned} &\int_0^\infty dT e^{-sT} \int_0^T dt_{\max} \int_{-\infty}^M dx_F G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|M-r, \theta) \\ &\quad \times G_M(M-r|x_F, t-t_{\max}-\theta) \\ &= G_M(M-\varepsilon|M-r, \theta) \tilde{G}_M(0|M-\varepsilon, s) \int_{-\infty}^M dx_F \tilde{G}_M(M-r|x_F, s) \\ &= \frac{2^{3/2} \varepsilon^2 r}{\sqrt{\pi} \theta^3 s} e^{-\frac{r^2}{2\theta}} e^{-\sqrt{2s}M} (1 - e^{-\sqrt{2s}r}). \end{aligned} \quad (4.5)$$

Après intégration sur la variable M puis inversion de Laplace, on trouve en utilisant la condition de normalisation (4.3)

$$\boxed{\text{--- Densité moyenne à partir du maximum à un temps fixé du brownien libre ---}} \\ \Xi(r, \theta, t) = \frac{4r\sqrt{t-|\theta|}\Phi^{(0)}\left(\frac{r}{\sqrt{|\theta|}}\right)}{|\theta|^{3/2}} \left(\frac{1}{\sqrt{2\pi}} - \Phi^{(0)}\left(\frac{r}{\sqrt{t-|\theta|}}\right) + \frac{r\Phi^{(1)}\left(\frac{r}{\sqrt{t-|\theta|}}\right)}{\sqrt{t-|\theta|}} \right), \quad (4.6)$$

où les fonctions $\Phi^{(0)}$ et $\Phi^{(1)}$ sont définies dans l'appendice A. L'intégration sur la variable θ (4.2), en utilisant encore une fois la transformée de Laplace, redonne la densité proche du maximum dans le cas du brownien libre (3.26)

$$\int_{-t}^t d\theta \Xi(r, \theta, t) = 8\sqrt{t} \left(\Phi^{(2)}\left(\frac{r}{\sqrt{t}}\right) - \Phi^{(2)}\left(2\frac{r}{\sqrt{t}}\right) \right) = \langle \rho_{\max}(r, t) \rangle. \quad (4.7)$$

L'intégration sur r donne la proportion de brownien libre tel que $t_{\max} + \theta \in [0, t]$

$$\int_0^\infty dr \Xi(r, \theta, t) = \frac{2}{\pi} \text{Arcsin} \left(\sqrt{1 - \frac{|\theta|}{t}} \right) = \int_{\max(-\theta, 0)}^{\min(t-\theta, t)} \frac{1}{\pi \sqrt{t_{\max}(t - t_{\max})}} dt_{\max}, \quad (4.8)$$

qui est conforme à la loi de l'arc sinus du temps d'atteinte du maximum pour le brownien libre [101].

4.1.2 Cas du pont

Dans le cas du pont, en modifiant l'équation (3.12), on trouve directement, dans le cas $\theta > 0$

$$\begin{aligned} \Xi_B(r, \theta, t) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{Z_B(\varepsilon)} \int_0^\infty dM \int_0^{t-\theta} dt_{\max} G_M(0|M-\varepsilon, t_{\max}) \\ &\quad \times G_M(M-\varepsilon|M-r, \theta) G_M(M-r|0, t-t_{\max}-\theta), \end{aligned} \quad (4.9)$$

avec $Z_B(\varepsilon)$ une constante de normalisation assurant (4.3). Le cas des $\theta < 0$ est identique à cause de la symétrie par inversion du temps du pont. L'intégrale sur la variable t_{\max} étant également un produit de convolution, il est plus simple de la calculer en utilisant la transformée de Laplace par rapport à la variable $T = t - \theta$

$$\begin{aligned} &\int_0^\infty dT e^{-sT} \int_0^T dt_{\max} G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|M-r, \theta) G_M(M-r|0, t-t_{\max}-\theta) \\ &= G_M(M-\varepsilon|M-r, \theta) \tilde{G}_M(0|M-\varepsilon, s) \tilde{G}_M(M-r|0, s) \\ &= \frac{2\varepsilon^2 r}{\sqrt{\pi s \theta^3}} e^{-\frac{r^2}{2\theta}} e^{-\sqrt{2s}M} \left(e^{-\sqrt{2s}|M-r|} - e^{-\sqrt{2s}(M+r)} \right), \end{aligned} \quad (4.10)$$

où l'on a utilisé le propagateur (3.8) et sa transformée de Laplace (3.9). Après intégration sur la variable M puis inversion de Laplace, on trouve en utilisant la condition de normalisation (4.3)

$$\boxed{\text{--- Densité moyenne à partir du maximum à un temps fixé du pont ---}} \\ \Xi_B(r, \theta, t) = \sqrt{\frac{2}{\pi}} \frac{r^2}{|\theta| \sqrt{|\theta|(1 - \frac{|\theta|}{t})}} \exp \left(-\frac{r^2}{2|\theta|(1 - \frac{|\theta|}{t})} \right). \quad (4.11)$$

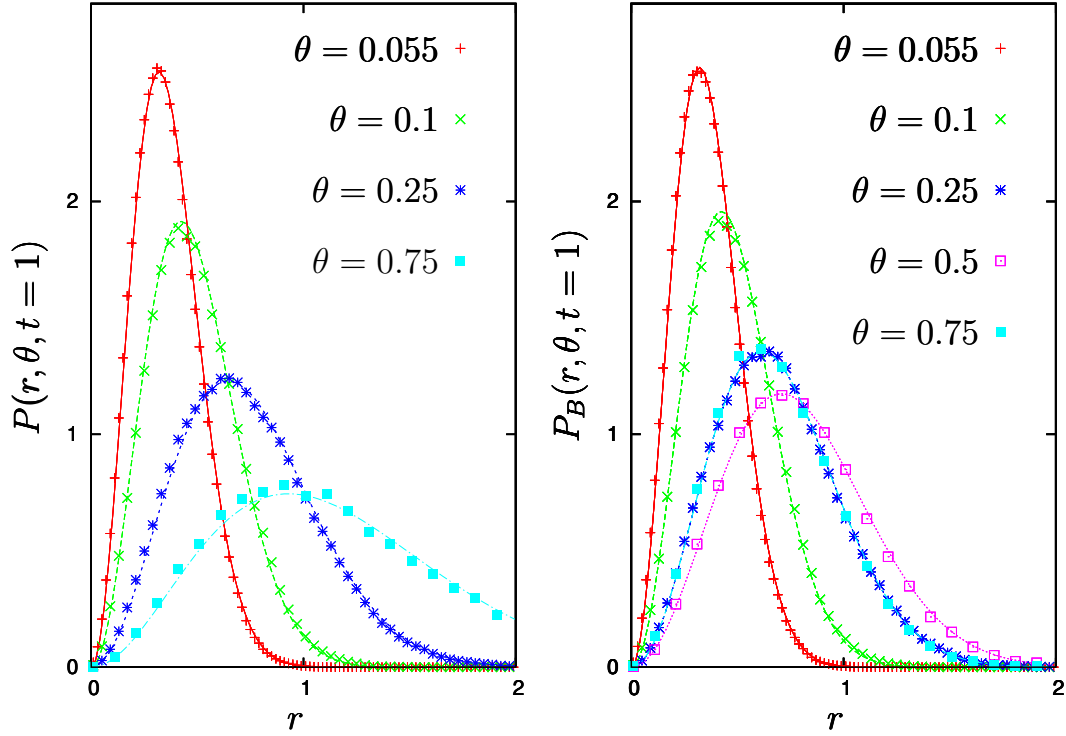


FIGURE 4.2: Comparaison de $P(r, \theta, t = 1)$ (4.15) obtenues en simulant 10^5 marches aléatoires de 10^4 pas et les formules analytiques (4.6) (4.8) pour le brownien libre et (4.11) (4.13) pour le pont. Les courbes sont tracées comme des fonctions de r pour différents θ fixés. Remarquons que la symétrie du pont impose $P_B(r, \theta, t) = P_B(r, t - \theta, t)$.

En intégrant sur la variable θ (4.2), en utilisant encore une fois la transformée de Laplace, on retrouve la densité proche du maximum dans le cas du pont (3.26)

$$\int_{-t}^t d\theta \Xi_B(r, \theta, t) = 4re^{-\frac{2r^2}{t}} = \langle \rho_{\max, B}(r, t) \rangle, \quad (4.12)$$

L'intégration sur r donne quant à elle

$$\int_0^\infty dr \Xi_B(r, \theta, t) = \left(1 - \frac{|\theta|}{t}\right), \quad (4.13)$$

ceci traduit la proportion de pont tel que $t_{\max} + \theta \in [0, t]$, t_{\max} étant une variable aléatoire de loi uniforme sur $[0, t]$ dans le cas du pont. Remarquons que dans le cas du pont à cause de la périodicité du pont, étudier la densité à r et θ du maximum sans considérer la condition $t_{\max} + \theta \in [0, t]$ (c'est-à-dire en renormalisant par 4.13), revient à étudier ce qui se passe en $t - \theta$ ainsi qu'en $-\theta$

$$\frac{\Xi_B(r, \theta, t)}{\int_0^\infty du \Xi_B(u, \theta, t)} = \frac{\Xi_B(r, t - \theta, t)}{\int_0^\infty du \Xi_B(u, t - \theta, t)}. \quad (4.14)$$

4.1.3 Comportements et résultats numériques

Il est en fait plus commode numériquement de travailler avec la quantité renormalisée que l'on interprète comme la probabilité d'être à une position r du maximum après un

temps θ fixé

$$P(r, \theta, t) = \frac{\Xi_B(r, \theta, t)}{\int_0^\infty du \Xi_B(u, \theta, t)} \quad (4.15)$$

$$\int_0^\infty dr P(r, \theta, t) = 1. \quad (4.16)$$

Nous avons tracé sur la figure 4.2 la comparaison entre les résultats obtenus via les formules (4.6) et (4.8) pour le brownien libre, (4.11) et (4.13) pour le pont avec des données obtenues en simulant 10^5 marches aléatoires de 10^4 pas. Il est intéressant de considérer la limite des θ et r petits pour analyser le voisinage immédiat du maximum. Comme on peut le voir sur la figure 4.2, les deux cas se ressemblent beaucoup dans cette limite. En effet, on peut réécrire

$$\begin{aligned} P_B(r, \theta, t) &= \frac{1}{\sqrt{|\theta|(1 - \frac{|\theta|}{t})}} \sqrt{\frac{2}{\pi}} \left(\frac{r}{\sqrt{|\theta|(1 - \frac{|\theta|}{t})}} \right)^2 \exp \left(-\frac{1}{2} \left(\frac{r}{\sqrt{|\theta|(1 - \frac{|\theta|}{t})}} \right)^2 \right) \\ &= \frac{1}{\sqrt{|\theta|(1 - \frac{|\theta|}{t})}} F \left(\frac{r}{\sqrt{|\theta|(1 - \frac{|\theta|}{t})}} \right), \quad F(x) = \sqrt{\frac{2}{\pi}} x^2 \exp \left(-\frac{x^2}{2} \right). \end{aligned} \quad (4.17)$$

En particulier, lorsque l'on considère la limite proche du maximum

$$\begin{cases} \theta \rightarrow 0 \\ r \rightarrow 0 \end{cases} \quad \text{avec } y = \frac{r}{\sqrt{|\theta|}} \text{ fixé}, \quad (4.18)$$

alors

$$\lim_{\theta \rightarrow 0} |\theta| P_B(y\sqrt{|\theta|}, \theta, t) = F(y). \quad (4.19)$$

La même limite appliquée au brownien libre donne le même résultat

$$|\theta| P(y\sqrt{|\theta|}, \theta, t) = F(y). \quad (4.20)$$

Cette propriété traduit encore plus précisément la réflexion que nous avons fait précédemment : le voisinage du maximum n'est pas sensible aux conditions aux bords. Il est possible de calculer cette quantité pour les autres mouvements browniens contraints. Le calcul se fait de la même manière en utilisant les propagateurs adéquats. Néanmoins, comme nous l'avons vu sur le cas du brownien libre (4.5) et du pont (4.10), ce calcul fait intervenir deux propagateurs dans le domaine de Laplace et un propagateur dans le domaine temporel, ce qui complique le calcul dans les autres cas. Nous ne nous attarderons pas plus sur cette étude temporelle.

4.2 Cas de plusieurs browniens

4.2.1 Plusieurs browniens indépendants

En utilisant ce formalisme des propagateurs, il est possible d'étudier la densité moyenne à partir du maximum pour n trajectoires indépendantes d'un même brownien par exemple dans le cas de n ponts indépendants.

$$\langle \rho_{\max, B, n}(r, t) \rangle = \left\langle \frac{1}{n} \sum_{i=1}^n \int_0^t \delta(x_{\max, n, B} - x_{i, B}(\tau) - r) d\tau \right\rangle, \quad (4.21)$$

où les trajectoires $x_{i,B}(\tau)$ sont n réalisations indépendantes entre elles de ponts et où le maximum est le maximum global sur l'ensemble des n ponts

$$x_{\max,n,B} = \max_{\tau \in [0,t], i \in [1,n]} x_{i,B}(\tau). \quad (4.22)$$

On a choisi de garder cette densité normalisée

$$\int_0^\infty dr \langle \rho_{\max,B,n}(r, t) \rangle = 1. \quad (4.23)$$

En utilisant la même méthode, on considère toutes les trajectoires qui passent à une position r de ce maximum global. Plusieurs cas se posent suivant si la trajectoire considérée est celle qui possède le maximum global ou non. Chacun de ces cas se décomposant en plusieurs contributions suivant l'ordre temporel des différents événements comme précédemment. En utilisant la propriété de Markov du mouvement brownien, on peut alors écrire chacune des contributions comme l'intégrale d'un produit de propagateur

$$\begin{aligned} \langle \rho_{\max,B,n}(r, t) \rangle &= \lim_{\varepsilon \rightarrow 0} \frac{1}{n Z_{B,n}(\varepsilon)} \left(\int_0^\infty dM \int_0^t dt_{\max} \right. \\ &\quad \int_0^{t_{\max}} d\tau G_M(0|M-r, \tau) G_M(M-r|M-\varepsilon, t_{\max}-\tau) G_M(M-\varepsilon|0, t-t_{\max}) G_M(0|0, t)^{n-1} \\ &\quad + \int_{t_{\max}}^t d\tau G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|M-r, \tau-t_{\max}) G_M(M-r|0, t-\tau) G_M(0|0, t)^{n-1} \\ &\quad + (n-1) \int_0^t d\tau G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|0, t-t_{\max}) G_M(0|M-r, \tau) \\ &\quad \left. \times G_M(M-r|0, t-\tau) G_M(0|0, t)^{n-2} \right), \end{aligned} \quad (4.24)$$

où la constante de normalisation est donnée par

$$Z_{B,n}(\varepsilon) = \int_0^\infty dM \int_0^t dt_{\max} G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|0, t-t_{\max}) G_M(0|0, t)^{n-1}, \quad (4.25)$$

et peut être calculée en utilisant la transformée de Laplace

$$Z_{B,n}(\varepsilon) \underset{\varepsilon \rightarrow 0}{\sim} \frac{2\varepsilon^2}{n(2\pi t)^{n/2}}. \quad (4.26)$$

En utilisant le domaine de Laplace, il est possible de calculer les intégrales temporelles, le calcul complet se trouve dans l'appendice B. Finalement, on trouve pour la densité à partir du maximum dans le cas de n ponts indépendants

— Densité moyenne à partir du maximum de n ponts indépendants —

$$\begin{aligned} \langle \rho_{\max,B,n}(r, 1) \rangle &= 4re^{-2r^2} \\ &\quad + \sum_{i=1}^{n-1} (-1)^i \sqrt{2\pi} \binom{n-1}{i} \left(\frac{e^{-2r^2} \operatorname{erf}(\sqrt{2}ir)}{\sqrt{i}} - \frac{e^{-2\frac{ir^2}{1+i}} \operatorname{erf}(\sqrt{2}\frac{ir}{\sqrt{1+i}})}{\sqrt{1+i}} \right), \end{aligned} \quad (4.27)$$

avec erf la fonction d'erreur (11). On retrouve dans le premier terme le premier moment de la densité à partir du maximum pour un pont $\langle \rho_{\max,B}(r, 1) \rangle$ qui correspond au cas où $n = 1$. Remarquons que le second terme est d'intégrale nulle, il est négatif pour des r

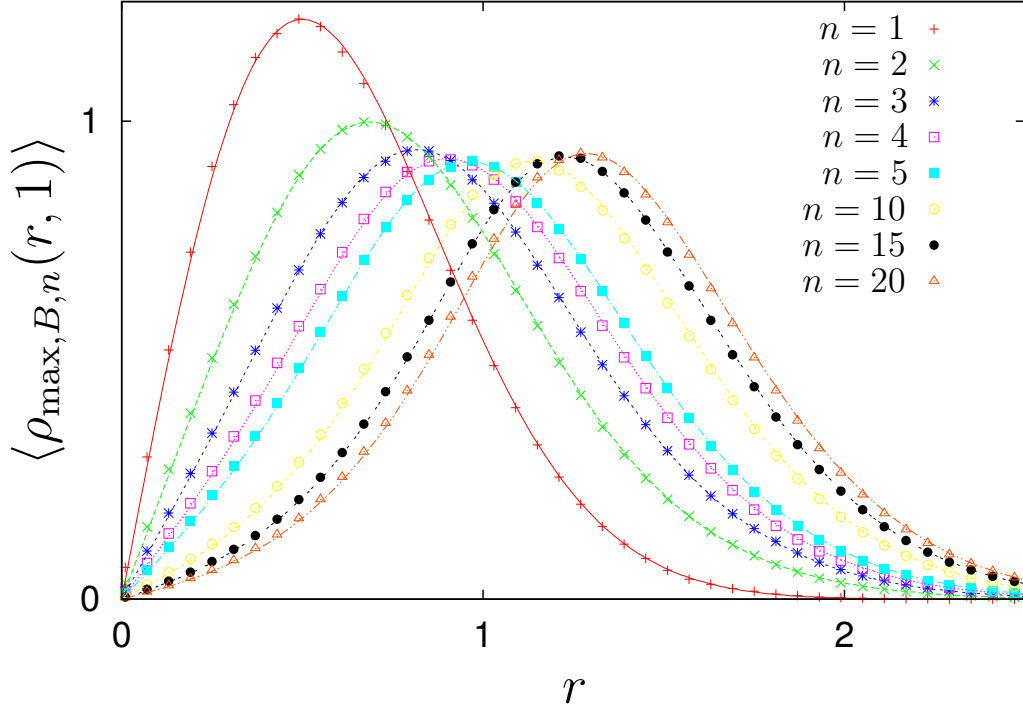


FIGURE 4.3: Densité à partir du maximum $\langle \rho_{\max, B, n}(r, 1) \rangle$ pour n browniens indépendants. La figure compare des données obtenues pour chacun des cas en moyennant sur 10^4 réalisations de n marches aléatoires de $N = 10^4$ pas avec la formule analytique (4.27).

petit puis positif à partir d'une certaine valeur de r (qui dépend de n). Il est possible à partir de cette densité de calculer la valeur moyenne de r

$$\langle r \rangle = \int_0^\infty dr \, r \langle \rho_{\max, B, n}(r, 1) \rangle = \frac{\sqrt{2\pi}}{4} \sum_{i=1}^n \frac{(-1)^{i+1}}{\sqrt{i}} \binom{n}{i}. \quad (4.28)$$

On peut interpréter ce résultat. Pour cela, notons que

$$\frac{\sqrt{2\pi}}{4\sqrt{i}} = \int_0^\infty dM e^{-2iM^2}. \quad (4.29)$$

On peut alors resommer la formule (4.28) en utilisant le binôme de Newton

$$\begin{aligned} \langle r \rangle &= \int_0^\infty dM \left(1 - \left(1 - e^{-2M^2} \right)^n \right) \\ &= \int_0^\infty dM \left(1 - p_{\max, B}(M, 1)^n \right) \\ &= \int_0^\infty dM \Pr. [x_{\max, n, B} > M] \\ &= \langle x_{\max, n, B} \rangle, \end{aligned} \quad (4.30)$$

où l'on a utilisé la statistique du maximum d'un seul pont $p_{\max,B}(x, t)$ donnée en équation (2.7). En effet, en réutilisant la définition (4.21)

$$\begin{aligned} \langle r \rangle &= \int_0^\infty dr \langle \rho_{\max,B,n}(r, 1) \rangle = \left\langle \frac{1}{n} \sum_{i=1}^n \int_0^1 d\tau \int_0^\infty dr \delta(x_{\max,n,B} - x_{i,B}(\tau) - r) r \right\rangle \\ &= \left\langle \int_0^1 d\tau (x_{\max,n,B} - x_{i,B}(\tau)) \right\rangle \\ &= \langle x_{\max,n,B} \rangle - \left\langle \int_0^1 d\tau x_{i,B}(\tau) \right\rangle, \end{aligned} \quad (4.31)$$

le second terme étant nul dans le cas du pont qui est symétrique. Il est possible d'étudier son comportement asymptotique dans la limite thermodynamique de grand n obtenue à partir d'un développement asymptotique de l'intégrale (4.30) comme nous le montrons dans l'appendice B

$$\langle r \rangle = \sqrt{\frac{\ln n}{2}} + \frac{\gamma_E}{2\sqrt{2 \ln n}} + o((\ln n)^{-3/2}), \quad (4.32)$$

où $\gamma_E = -\int_0^\infty e^{-t} \ln t dt = 0.577\dots$ est la constante gamma d'Euler. Ce résultat montre que typiquement, le pont qui possède le maximum $x_{\max,n,B}$ est très loin de la très grande majorité des autres ponts qui restent proches de l'origine. Cette interprétation peut être rendue plus précise en étudiant directement le comportement de $\langle \rho_{\max,B,n}(r = \sqrt{\frac{\ln n}{2}} + z, 1) \rangle$ lorsque le nombre de ponts n tend vers l'infini. Dans cette limite, la contribution principale est donnée par la dernière intégrale (4.24) où le maximum global $x_{\max,n,B}$ et le point à la position $x_{\max,n,B} - r$ n'appartiennent pas au même pont. Dans cette limite où $r = \sqrt{\frac{\ln n}{2}} + z$ tend également vers l'infini, ces deux événements sont essentiellement indépendants. La distribution du maximum d'un pont étant donnée par la loi (2.7), la loi du maximum global, qui est le maximum des n maxima des ponts indépendants, est une loi de Gumbel dont les comportements asymptotiques de la moyenne $\sim a_n$ et de l'écart type $\sim b_n$ peuvent être calculés en utilisant la théorie rappelée au chapitre 1

$$a_n = \sqrt{\frac{\ln n}{2}}, \quad (4.33)$$

$$b_n = \frac{1}{2\sqrt{2 \ln n}}. \quad (4.34)$$

Dans la limite que l'on considère, les fluctuations du maximum global sont imperceptibles $b_n \rightarrow 0$ et sa position est fixée à a_n . Ainsi, en reprenant la formule (4.21)

$$\langle \rho_{\max,B,n}(a_n + z, t) \rangle \underset{n \rightarrow \infty}{\approx} \left\langle \frac{1}{n} \sum_{i=1}^n \int_0^t \delta(a_n - x_{i,B}(\tau) - a_n - z) d\tau \right\rangle, \quad (4.35)$$

Or la contribution de chacun des ponts est identique et est le temps local $\langle T_{loc,B}(z, 1) \rangle$ c'est-à-dire la densité. Finalement, la densité à partir du maximum pour un grand nombre de ponts est donnée par

— Densité à partir du maximum d'un grand nombre de ponts indépendants —

$$\langle \rho_{\max,B,n}(\sqrt{\frac{\log(n)}{2}} + z, 1) \rangle \underset{n \rightarrow \infty}{\sim} \langle T_{loc,B}(z, 1) \rangle = \sqrt{\frac{\pi}{2}} \operatorname{erfc}(\sqrt{2}|z|). \quad (4.36)$$

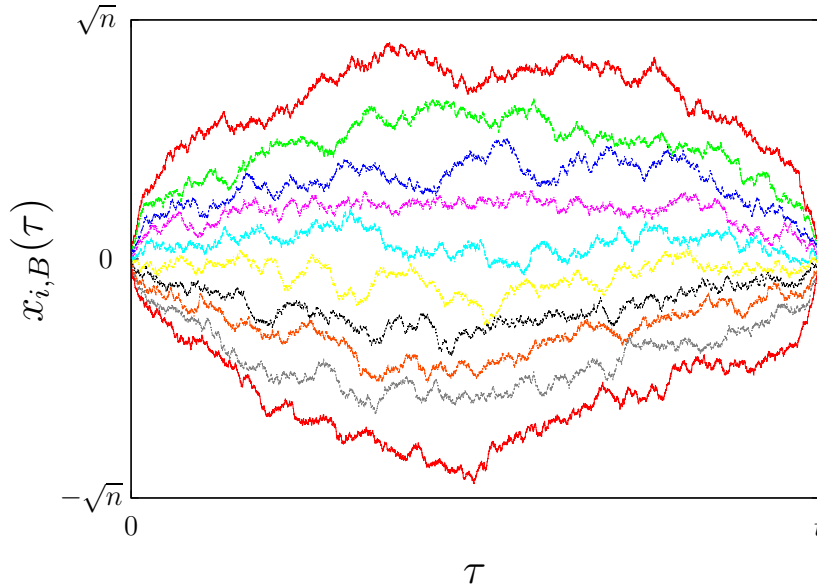


FIGURE 4.4: Exemple de $n = 10$ ponts répulsifs c'est-à-dire contraints à ne pas se croiser. Ces trajectoires sont générées à partir du mouvement brownien de Dyson [46] décrit en appendice C.

Cette formule ressemble au cas des variables indépendantes à décroissance plus rapide qu'une exponentielle (1.19) puisque les ponts sont décorrélés entre eux. Nous avons démontré cette formule par des arguments heuristiques, le calcul direct, reproduit en appendice B, confirme ce résultat. Remarquons enfin que lorsque l'on s'intéresse au régime de bord à r petit, on peut montrer à partir de la formule (4.27)

$$\lim_{n \rightarrow \infty} n^2 \langle \rho_{\max, B, n} \left(\frac{x}{n}, 1 \right) \rangle = 4x. \quad (4.37)$$

Ce régime correspond aux événements où le maximum global $x_{\max, n, B}$ et le point à la position $x_{\max, n, B} - r$ appartiennent à un même pont. On retrouve ainsi le même comportement que tous les cas précédents, comme le montre le tableau 3.1, puisqu'il ne dépend pas des conditions aux bords.

4.2.2 Ponts répulsifs

Il est également possible de considérer n ponts contraints à ne pas se croiser $x_{1,B}(\tau) < x_{2,B}(\tau) < \dots < x_{n,B}(\tau)$ pour tous les temps $\tau \in [0, 1]$ comme le montre la figure 4.4. De telles trajectoires sont appelées marcheurs répulsifs - *vicious walkers*. Introduit pour l'étude des polymères [36], ce modèle est désormais populaire dans plusieurs domaines de la physique [17, 54, 60] grâce à sa connexion avec les matrices aléatoires, les théories de jauge de type Yang-Mills sur la sphère [67, 144], ainsi que les modèles dans la classe d'universalité de Kardar-Parisi-Zhang [94] en particulier la distribution du maximum global de ces trajectoires est après un changement d'échelle le processus d'Airy₂ soustrait d'une parabole typique de cette classe d'universalité [141]. L'étude de ces trajectoires contraintes se réalise en considérant le problème quantique associé sur des fermions, le principe de Fermi interdisant aux fermions de se croiser. Les fonctions propres sont alors construites à l'aide du déterminant de Slater qui impose la condition de non-croisement. Il reste néanmoins difficile d'obtenir avec cette méthode une formule explicite pour $\langle \rho_{\max, B, n}^{\text{Vicious}}(r, t = 1) \rangle$,

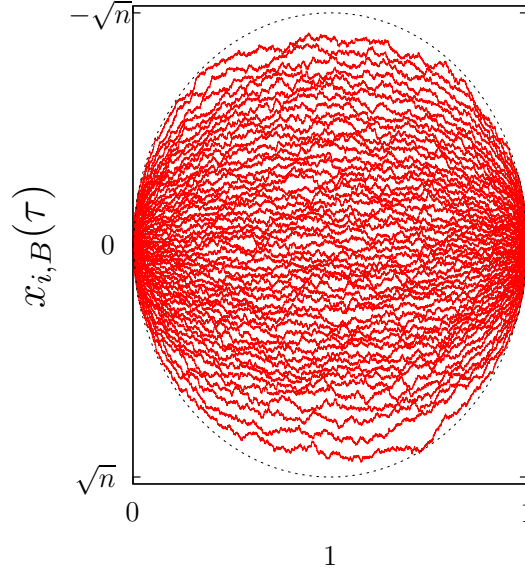


FIGURE 4.5: Dans la limite d'un grand nombre de ponts répulsifs, ici $n = 50$, la trajectoire la plus haute et la trajectoire la plus basse sont, à des fluctuations près, déterministes et sont des demi-cercles $\pm 2\sqrt{n\tau(1-\tau)}$ (en pointillés).

densité à partir du maximum de n ponts contraints à ne pas se croiser. Néanmoins, il reste possible d'étudier cette quantité dans la limite de grand n grâce à la correspondance de ce modèle avec les matrices aléatoires. En effet, on peut montrer en utilisant les propagateurs construits à partir de ces fonctions d'ondes de Slater [140] que la loi jointe des différentes positions à un temps τ fixée est

$$P_{\text{joint}}(x_{1,B}(\tau), \dots, x_{n,B}(\tau)) = \frac{1}{Z_{B,n}(\tau)} \prod_{i < j} (x_{i,B}(\tau) - x_{j,B}(\tau))^2 \exp \left(- \sum_{i=1}^n \frac{x_{i,B}(\tau)^2}{2\tau(1-\tau)} \right), \quad (4.38)$$

où $Z_{B,n}(\tau)$ est une constante de normalisation qui peut être évaluée à partir d'une intégrale de Selberg [113]. Cette loi est la même que la loi jointe des n valeurs propres λ_i du modèle des matrices aléatoires gaussiennes unitaires (que nous rediscuterons dans la partie suivante) avec la correspondance $\lambda_i \sqrt{2\tau(1-\tau)} = x_{i,B}$. En particulier, le chemin le plus haut $x_{n,B}(\tau)$ correspond à la même statistique que la valeur propre maximale du modèle de matrice. En utilisant les résultats de Tracy et Widom [153, 154] sur la statistique de cette valeur propre maximale, on trouve, à τ fixé

$$\frac{x_{n,B}(\tau)}{\sqrt{2\tau(1-\tau)}} = \sqrt{2n} + O(n^{-1/6}), \quad (4.39)$$

où les fluctuations $O(n^{-1/6})$ sont gouvernées par la distribution de Tracy-Widom qui sera discutée dans la partie suivante. De la même façon, le chemin le plus bas $x_{1,B}(\tau)$ est, à des fluctuations près, $x_{1,B}(\tau) = -2\sqrt{\tau(1-\tau)n}$. Un exemple de trajectoires est montré dans la figure 4.5 où l'on a tracé 50 ponts répulsifs. En particulier, le maximum global des ponts répulsifs est

$$x_{\max,n,B}^{\text{Vicious}} = \max_{\tau \in [0,t], i \in [1,n]} x_{i,B}(\tau) = \sqrt{n} + O(n^{-1/6}), \quad (4.40)$$

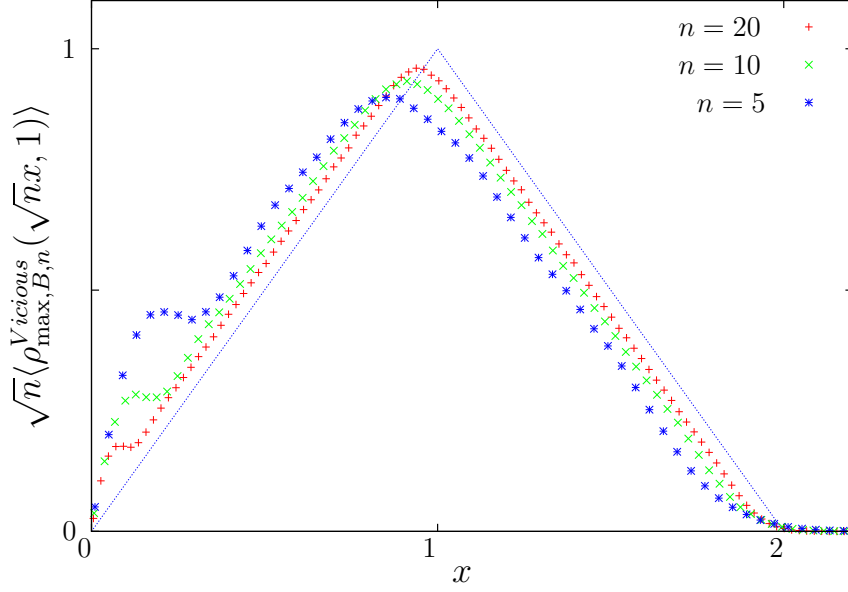


FIGURE 4.6: Résultats numériques obtenus en simulant des marches aléatoires de 10^4 pas contraintes à ne pas se croiser démarrant et terminant à l'origine (moyennés 10^4 fois) pour différents nombres de marches n : les points bleus sont obtenus pour $n = 5$ ponts, les points verts pour $n = 10$ ponts et les points rouges pour $n = 20$ ponts. La ligne pointillée bleue représente le comportement asymptotique (4.45) dans la limite d'un grand nombre n de marcheurs.

et est atteint par la plus haute trajectoire $x_n(\tau)$ en $\tau = 1/2$, à des fluctuations près. Plus précisément, il a été montré par Prähofer et Spohn [131, 132] que le régime d'échelle autour de ce maximum global est gouverné par le processus d'Airy₂ $\mathcal{A}_2(u)$

$$n^{1/6} \left(x_{n,B} \left(\tau = \frac{1}{2} + \frac{u}{2n^{1/3}} \right) - \sqrt{n} \right) = \mathcal{A}_2(u) - u^2. \quad (4.41)$$

Nous nous bornerons ici à l'étude de la densité à partir du maximum dans un régime où ces fluctuations sont sous-dominantes et l'approximation (4.40) sera suffisante (régime dit de volume). Nous rediscuterons brièvement (et numériquement) le régime où ces fluctuations interviennent, le régime de bord, à la fin de cette section. On définit la densité à partir du maximum dans le cas de n ponts répulsifs comme nous l'avons fait pour les ponts indépendants (4.21)

$$\langle \rho_{\max, B, n}^{Vicious}(r, t = 1) \rangle = \left\langle \frac{1}{n} \sum_{i=1}^n \int_0^1 \delta(x_{\max, n, B}^{Vicious} - x_{i,B}(\tau) - r) d\tau \right\rangle. \quad (4.42)$$

Même s'il est possible d'étudier précisément la distribution du maximum des n ponts répulsifs [137], il est difficile d'extraire une formule analytique pour la densité à partir du maximum (4.42) à n fini. Néanmoins, dans la limite de grand n , il est possible à partir du résultat (4.40) de discuter le comportement de cette densité. En effet, au premier ordre, les fluctuations du maximum sont négligeables et on peut remplacer $x_{\max, n, B}^{Vicious}$ par $\sqrt{2}$ directement dans la définition (4.42) de la densité à partir du maximum qui se réduit dans ce cas au temps local translaté

$$\langle \rho_{\max, B, n}^{Vicious}(r, 1) \rangle \underset{n \rightarrow \infty}{=} \left\langle \frac{1}{n} \sum_{i=1}^n \int_0^1 \delta(\sqrt{2} - r - x_{i,B}(\tau)) d\tau \right\rangle \quad (4.43)$$

$$= \langle T_{loc, B, n}^{Vicious}(r - \sqrt{2}, 1) \rangle. \quad (4.44)$$

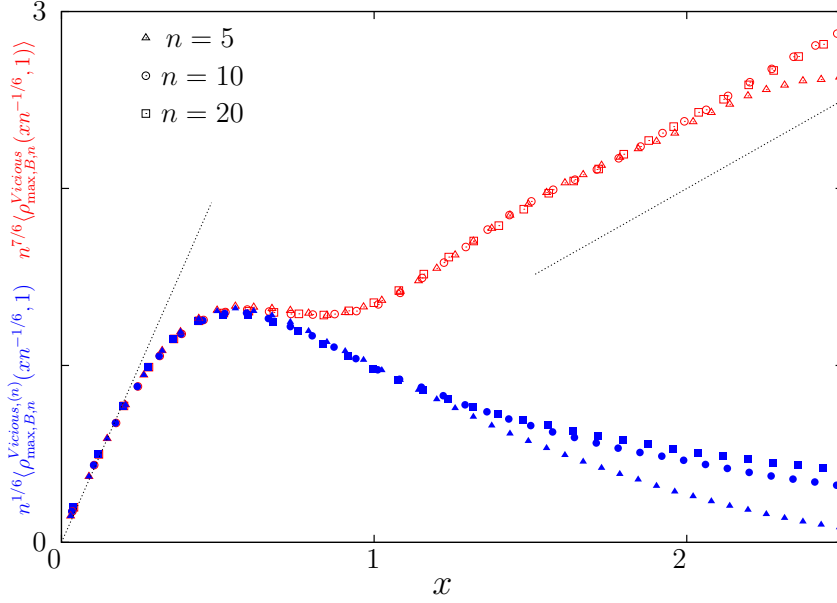


FIGURE 4.7: Densité à partir du maximum dans le régime de bord (4.46) en rouge et densité à partir du maximum du plus haut pont uniquement typique (4.49) en bleu pour différents nombres de ponts : 5 ponts pour les triangles, 10 ponts pour les ronds et 20 ponts pour les carrés. Ces quantités convergent respectivement vers les fonctions ρ_{edge} et ρ_{∞} qui possèdent bien les comportements asymptotiques discutés (4.47) et (4.49). Ces données sont obtenues en générant des marches aléatoires de 10^4 pas contraintes à ne pas se croiser démarrant et terminant à l'origine et moyennées 10^4 fois.

Il est alors possible d'étudier ce temps local en utilisant la correspondance avec la théorie des matrices aléatoires. Cette étude est faite en appendice C. On montre alors que la densité à partir du maximum suit la loi d'échelle dans le régime de volume

Densité à partir du maximum pour un grand nombre de ponts répulsifs

$$\sqrt{n} \langle \rho_{\max, B, n}^{Vicious}(\sqrt{n}x, 1) \rangle \underset{n \rightarrow \infty}{\sim} \begin{cases} x & \text{pour } x < 1 \\ 2 - x & \text{pour } 1 < x < 2 \\ 0 & \text{pour } 2 < x. \end{cases} \quad (4.45)$$

Comme on peut le voir sur la figure 4.6, ce comportement est bien observé dans la limite d'un grand nombre de ponts. Remarquons sur cette figure qu'il existe un régime différent pour x petit dû aux contributions du pont le plus haut (qui est donc celui qui atteint le maximum $x_{\max, n, B}^{Vicious}$) correspondant au régime de bord. Ce régime possède le scaling $n^{7/6} \langle \rho_{\max, B, n}^{Vicious}(xn^{-1/6}, 1) \rangle$ qui est le régime des fluctuations du maximum (4.41). On s'attend ainsi

$$\lim_{n \rightarrow \infty} n^{7/6} \langle \rho_{\max, B, n}^{Vicious}(xn^{-1/6}, 1) \rangle = \rho_{edge}(x), \quad (4.46)$$

où les comportements asymptotiques de cette fonction d'échelle sont imposés à x petit car le comportement de la densité proche du maximum pour le brownien ne dépend pas des contraintes et à grand x pour venir se raccorder avec le régime de volume (4.45)

$$\rho_{edge}(x) = \begin{cases} 4x & \text{pour } x \ll 1 \\ x & \text{pour } x \gg 1. \end{cases} \quad (4.47)$$

Ces comportements sont bien vérifiés numériquement comme on peut le voir sur la figure 4.7. Il est également possible d'étudier la contribution complète du brownien le plus haut $x_{n,B}$ uniquement

$$\langle \rho_{\max,B,n}^{Vicious,(n)}(r, 1) \rangle = \langle \int_0^1 \delta(x_{\max,n,B}^{Vicious} - x_{n,B}(\tau) - r) d\tau \rangle, \quad (4.48)$$

dont il est possible de donner le comportement typique ainsi que son comportement à très petit argument également tracé en figure 4.7

$$\lim_{n \rightarrow \infty} n^{1/6} \langle \rho_{\max,B,n}^{Vicious,(n)}(xn^{-1/6}, 1) \rangle = \rho_{\infty}(x), \quad (4.49)$$

$$\rho_{\infty}(x) \underset{x \rightarrow 0}{\sim} 4x. \quad (4.50)$$

L'étude complète de cette quantité étant difficile. Mentionnons également pour terminer que l'étude d'excursions répulsives est également riche et est reliée au modèle des matrices aléatoires de Wishart [140, 156].

Chapitre 5

La méthode de l'intégrale de chemin

Même s'il est facile de calculer par la méthode des propagateurs la densité moyenne à partir du maximum $\langle \rho_{\max}(r, t) \rangle$, il peut être difficile d'étudier les différents moments pour les différents mouvements browniens ou encore d'obtenir la loi complète. Pour cette étude, il est plus commode d'utiliser le formalisme de l'intégrale de chemin. Nous généralisons ici notre problème en essayant de calculer les statistiques de n'importe quelles fonctionnelles du maximum du mouvement brownien $\mathcal{O}_{\max}(t) = \int_0^t d\tau V(x_{\max} - x(\tau))$ où $V(x)$ est un potentiel quelconque, en particulier, on retrouve la densité à partir du maximum avec le potentiel $V(x) = \delta(x)$. Pour étudier cette statistique, il est utile de considérer sa transformée de Laplace. Dans cette section, nous présentons donc une manière de calculer cette transformée de Laplace pour n'importe quel potentiel $V(x)$ en utilisant des techniques d'intégrales de chemin appliquées au brownien libre (section 5.1.1) et au pont (5.1.2). Nous étudierons alors la statistique complète de la densité à partir du maximum $\rho_{\max}(r, t)$ en section 5.2.1. Nous appliquerons également nos résultats à d'autres problèmes dans les parties 5.2.2 et 5.2.3.

5.1 Principe

5.1.1 Brownien libre

Nous cherchons à calculer la transformée de Laplace $\langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle$ où le potentiel $V(x)$ est quelconque et où x_{\max} est la position maximale du brownien libre $x(\tau)$ sur $\tau \in [0, t]$. Dans ce but, nous découpons tout d'abord l'intervalle temporel $[0, t]$ en deux parties $[0, t_{\max}]$ et $[t_{\max}, t]$ avec t_{\max} le temps où le maximum x_{\max} est atteint. Pour un brownien libre, la densité de probabilité de t_{\max} est donné par la loi de l'arc sinus [101]

$$P(t_{\max}) = \frac{1}{\pi \sqrt{t_{\max}(t - t_{\max})}} . \quad (5.1)$$

Grâce à la propriété de Markov du mouvement brownien, les deux parties sont indépendantes. Dans chacune des parties, le processus $y(\tau) = x_{\max} - x(\tau)$ est un mouvement brownien qui reste positif sur tout l'intervalle. En inversant en plus le sens du temps dans la partie $[0, t_{\max}]$, on remarque que la trajectoire $y(\tau)$ est constituée de deux méandres indépendants de durées respectives t_{\max} et $t - t_{\max}$, on parle dans ce cas de double méandre [88] (cf. la figure 5.1). Ainsi, grâce à la propriété de Markov du brownien, on peut écrire

$$\langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \int_0^t dt_{\max} \varphi(t_{\max}) \varphi(t - t_{\max}) \quad (5.2)$$

$$\varphi(\tau) = \frac{1}{\sqrt{\pi \tau}} \langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_{Me} , \quad (5.3)$$

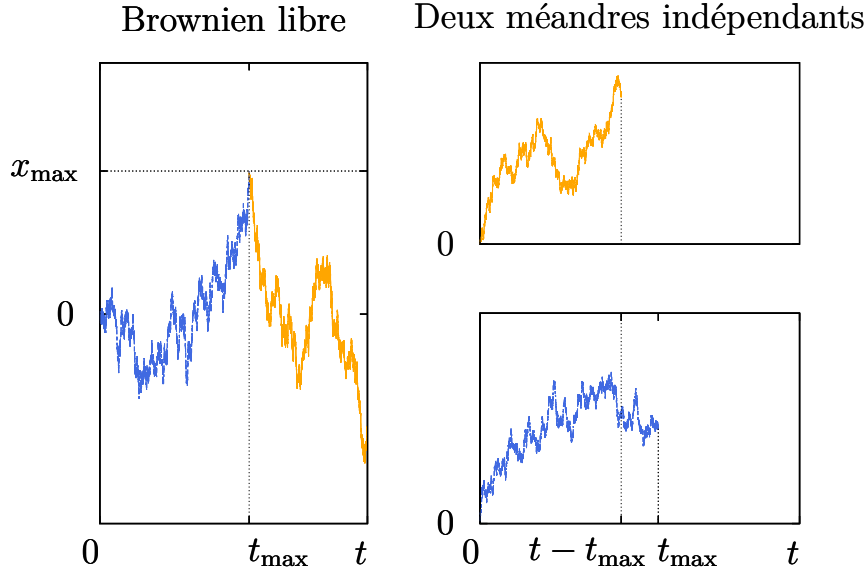


FIGURE 5.1: Un brownien libre peut être vu comme l'assemblage de deux méandres indépendants de durée respective t_{\max} et $t - t_{\max}$ où t_{\max} est distribué suivant la loi de l'arc sinus (5.1). On parle dans ce cas de double méandre [88].

où $\langle \cdots \rangle_{Me}$ est désormais une moyenne sur les trajectoires $y(\tau)$ du méandre. Dans (5.3) le terme $1/\sqrt{\pi\tau}$ provient de la loi de l'arc sinus due à t_{\max} (5.1). Il est alors possible de calculer la fonctionnelle $\varphi(\tau)$ du méandre à l'aide de techniques d'intégrales de chemin [103]. Néanmoins, la structure de convolution de l'équation (5.2) suggère plutôt de calculer sa transformée de Laplace par rapport à t

$$\int_0^\infty e^{-st} \langle e^{-\lambda \int_0^t V(x_{\max} - x(\tau)) d\tau} \rangle dt = \tilde{\varphi}(s)^2, \quad (5.4)$$

$$\tilde{\varphi}(s) = \int_0^\infty e^{-st} \varphi(t) dt. \quad (5.5)$$

Encore une fois, il est nécessaire d'introduire une régularisation car il n'est pas possible d'imposer au méandre simultanément $y(0) = 0$ et $y(0^+) > 0$. Pour éviter cette difficulté [105], on régularise en imposant à la place $y(0) = \varepsilon$, et l'on prendra la limite $\varepsilon \rightarrow 0$ à la fin du calcul. La formule de Feynman-Kac donne alors

$$\langle e^{-\lambda \int_0^\tau V(y(u)) du} \rangle_{Me} = \lim_{\varepsilon \rightarrow 0} \frac{\int_0^\infty \langle y_F | e^{-H_\lambda \tau} | \varepsilon \rangle dy_F}{\int_0^\infty \langle y_F | e^{-H_0 \tau} | \varepsilon \rangle dy_F}, \quad (5.6)$$

où le hamiltonien est donné par

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dy^2} + \lambda V(y) + V_{\text{wall}}(y), \quad (5.7)$$

avec $V_{\text{wall}}(y)$ une barrière de potentiel qui oblige la particule à rester du côté $y \geq 0$: $V_{\text{wall}}(y) = 0$ pour $y \geq 0$ et $V_{\text{wall}}(y) = +\infty$ pour $y < 0$. Il est facile de calculer les fonctions propres ψ_k du hamiltonien $H_0 = -\frac{1}{2} \frac{d^2}{dy^2} + V_{\text{wall}}(y)$:

$$\psi_k(y) = \sqrt{\frac{2}{\pi}} \sin(ky) \quad \text{d'énergie} \quad \frac{k^2}{2}, \quad (5.8)$$

ainsi, le dénominateur de (5.6) se calcule en utilisant

$$\langle y_F | e^{-H_0 \tau} | \varepsilon \rangle = \frac{2}{\pi} \int_0^\infty dk \sin(k y_F) \sin(k \varepsilon) e^{-\tau \frac{k^2}{2}} \underset{\varepsilon \rightarrow 0}{\sim} \frac{2 y_F \varepsilon}{\sqrt{2 \pi \tau}^{3/2}} e^{-\frac{y_F^2}{2 \tau}}, \quad (5.9)$$

ce qui donne directement

$$\int_0^\infty dy_F \langle y_F | e^{-H_0 \tau} | \varepsilon \rangle \underset{\varepsilon \rightarrow 0}{\sim} \frac{2 \varepsilon}{\sqrt{2 \pi \tau}}. \quad (5.10)$$

En notant Ψ_n les fonctions propres du hamiltonien H_λ ainsi que E_n ces énergies propres associées (nous supposons ici le spectre discret, néanmoins le calcul se généralise sans difficulté au cas où le spectre est continu), nous pouvons alors réécrire

$$\langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_{Me} = \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{2 \pi \tau}}{2 \varepsilon} \int_0^\infty dy_F \sum_{n=0}^\infty \Psi_n^*(y_F) \Psi_n(\varepsilon) e^{-E_n \tau}. \quad (5.11)$$

En utilisant la condition $\Psi_n(0) = 0$ imposée par la barrière de potentiel infinie en $x = 0$, on a $\lim_{\varepsilon \rightarrow 0} \frac{\Psi_n(\varepsilon)}{\varepsilon} = \Psi_n'(0)$ et donc

$$\langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_{Me} = \sqrt{\frac{\pi \tau}{2}} \int_0^\infty dy_F \sum_{n=0}^\infty \Psi_n^*(y_F) \Psi_n'(0) e^{-E_n \tau}. \quad (5.12)$$

En notant que le propagateur, qu'on décide dans cette partie de noter $G_\tau(x, y)$, s'écrit

$$G_\tau(x, y) = \langle y | e^{-H_\lambda \tau} | x \rangle = \sum_{n=0}^\infty \Psi_n^*(y) \Psi_n(x) e^{-E_n \tau}, \quad (5.13)$$

on trouve finalement

$$\langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_{Me} = \sqrt{\frac{\pi \tau}{2}} \int_0^\infty dy_F \partial_x G_\tau(0, y_F). \quad (5.14)$$

Il est alors possible d'utiliser la formule (5.14) pour calculer la transformée de Laplace $\tilde{\varphi}(s)$ définie en (5.5) et utile pour (5.4)

$$\begin{aligned} \tilde{\varphi}(s) &= \int_0^\infty e^{-s \tau} \varphi(\tau) d\tau = \frac{1}{\sqrt{2}} \int_0^\infty e^{-s \tau} \int_0^\infty dy_F \partial_x G_\tau(0, y_F) \\ &= \frac{1}{\sqrt{2}} \int_0^\infty dy_F \sum_{n=0}^\infty \frac{\Psi_n^*(y_F) \Psi_n'(0)}{s + E_n} = \frac{1}{\sqrt{2}} \int_0^\infty dy_F \partial_x \tilde{G}_s(0, y_F), \end{aligned} \quad (5.15)$$

où on a noté $\tilde{G}_s(x, y)$ la transformée de Laplace de $G_t(x, y)$ par rapport à la variable t

$$\tilde{G}_s(x, y) = \sum_{n=0}^\infty \frac{\Psi_n^*(y) \Psi_n(x)}{s + E_n}. \quad (5.16)$$

De plus, la fonction $\tilde{G}_s(x, y)$ est la fonction de Green satisfaisant

$$[H_\lambda + s] \tilde{G}_s(x, y) = \delta(x - y), \quad (5.17)$$

avec les conditions $\tilde{G}_s(x, 0) = \tilde{G}_s(0, y) = 0$. Pour calculer cette fonction de Green $\tilde{G}_s(x, y)$, on cherche deux solutions indépendantes $u_s(y)$ et $v_s(y)$ de l'équation homogène associée

$$[H_\lambda + s] \psi(y) = 0 \quad (5.18)$$

avec les conditions aux limites $u_s(0) = 0$ et $\lim_{y \rightarrow \infty} v_s(y) = 0$. À l'aide de ces deux solutions $u_s(y)$ et $v_s(y)$, on peut calculer $\tilde{G}_s(x, y)$ [97]

$$\tilde{G}_s(x, y) = \begin{cases} \frac{2}{W} u_s(x) v_s(y) & \text{pour } x \leq y, \\ \frac{2}{W} u_s(y) v_s(x) & \text{pour } x \geq y, \end{cases} \quad (5.19)$$

avec W le wronskien associé aux fonctions $\{u_s(y), v_s(y)\}$

$$W = u'_s(y)v_s(y) - u_s(y)v'_s(y). \quad (5.20)$$

Ce wronskien est ici indépendant de y car l'équation différentielle ne fait pas intervenir de dérivée première. On utilisant alors tous ces résultats, on obtient

$$\tilde{\varphi}(s) = \int_0^\infty e^{-s\tau} \varphi(\tau) d\tau = \frac{\sqrt{2}}{W} \int_0^\infty dy_F u'_s(0) v_s(y_F). \quad (5.21)$$

Cette formule (5.21) associée aux résultats (5.4) et (5.5) permet de calculer la transformée de Laplace $\langle e^{-\lambda \int_0^t V(x_{\max} - x(\tau)) d\tau} \rangle$ pour n'importe quel potentiel $V(x)$. Pour obtenir des résultats explicites, il suffit alors d'étudier le hamiltonien (5.7) et de résoudre l'équation homogène associée (5.18) ce que nous ferons sur quelques exemples dans la section (5.2).

Récapitulatif : cas du brownien libre

On cherche deux solutions indépendantes de l'équation homogène $\{u_s(y), v_s(y)\}$

$$\left[-\frac{1}{2} \frac{d^2}{dy^2} + \lambda V(y) + s \right] \psi(y) = 0,$$

vérifiant les conditions aux limites

$$u_s(0) = 0, \quad \lim_{y \rightarrow \infty} v_s(y) = 0.$$

En notant le wronskien $W = u'_s(y)v_s(y) - u_s(y)v'_s(y)$, la double transformée de Laplace recherchée est

$$\int_0^\infty e^{-st} \langle e^{-\lambda \int_0^t V(x_{\max} - x(\tau)) d\tau} \rangle dt = \left(\frac{\sqrt{2}}{W} \int_0^\infty dy_F u'_s(0) v_s(y_F) \right)^2. \quad (5.22)$$

5.1.2 Pont

Il est également possible d'étudier directement les transformées de Laplace des fonctionnelles du maximum pour un pont $\langle e^{-\lambda \int_0^t V(x_{\max, B} - x_B(\tau)) d\tau} \rangle$ où $x_{\max, B}$ est la position maximale du pont $x_B(\tau)$ pour $\tau \in [0, t]$. On peut alors utiliser la même logique que précédemment en découpant l'intervalle de temps $[0, t]$ en deux parties $[0, t_{\max}]$ et $[t_{\max}, t]$ avec t_{\max} le temps où le pont atteint son maximum, la variable t_{\max} étant cette fois uniforme $P(t_{\max}) = 1/t$ à cause des conditions aux limites périodiques du pont. Il existe en réalité une méthode plus simple qui utilise la construction de Vervaat que l'on a introduit dans

la section 3.3 qui transforme les fonctionnelles du maximum du pont en fonctionnelles de l'excursion qui ont déjà été étudiées [103, 105, 106] (cf. figure 3.2). En effet, on peut généraliser la relation (3.20) à toutes les fonctionnelles

$$\int_0^t V(x_{\max,B} - x_B(\tau))d\tau \stackrel{\text{loi}}{=} \int_0^t V(x_{\max,E} - x_E(\tau))d\tau \stackrel{\text{loi}}{=} \int_0^t V(x_E(\tau))d\tau . \quad (5.23)$$

Encore une fois, il est nécessaire d'introduire une régularisation car il n'est pas possible d'imposer pour l'excursion simultanément $x_E(0) = x_E(t) = 0$ et $x_E(0^+) > 0$ ainsi que $x_E(t^-) > 0$. Pour éviter cette difficulté [105], on régularise en imposant à la place $x_E(0) = x_E(t) = \varepsilon$, on prendra la limite $\varepsilon \rightarrow 0$ à la fin du calcul. Il est alors possible d'étudier la transformée de Laplace de la fonctionnelle grâce à la formule de Feynmann-Kac

$$\langle e^{-\lambda \int_0^t V(x_{\max,B} - x_B(\tau))d\tau} \rangle = \lim_{\varepsilon \rightarrow 0} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle} , \quad (5.24)$$

où H_λ est le même hamiltonien que précédemment. On peut facilement calculer le dénominateur de (5.24) en utilisant les résultats (5.9)

$$\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle \underset{\varepsilon \rightarrow 0}{\sim} \varepsilon^2 \sqrt{\frac{2}{\pi t^3}} . \quad (5.25)$$

La formule (5.24) est générale et permet de calculer la statistique des fonctionnelles du maximum du pont pour n'importe quel potentiel $V(x)$.

— Récapitulatif : cas du pont —

On cherche les fonctions et énergies propres du hamiltonien

$$H_\lambda = \begin{cases} +\infty & y < 0 , \\ -\frac{1}{2} \frac{d^2}{dy^2} + \lambda V(y) & y \geq 0 , \end{cases}$$

permettant directement de calculer la transformée de Laplace

$$\langle e^{-\lambda \int_0^t V(x_{\max,B} - x_B(\tau))d\tau} \rangle = \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{2\pi} t^{3/2}}{2\varepsilon^2} \langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle . \quad (5.26)$$

5.2 Applications à l'étude de certaines fonctionnelles du maximum

Dans cette section, nous allons utiliser le formalisme précédent pour étudier plusieurs fonctionnelles à la fois pour le brownien libre et pour le pont. Nous étudierons tout d'abord la statistique complète de la densité à partir du maximum dans le paragraphe 5.2.1 puis la famille de fonctionnelles correspondante aux potentiels $V(x) = x^\alpha$ dont nous discuterons les applications dans le paragraphe 5.2.2.

5.2.1 Statistique complète de la densité à partir du maximum

Brownien libre

Pour calculer la statistique complète de la densité à partir du maximum $\rho_{\max}(r, t) = \int_0^t \delta(x_{\max} - x(\tau) - r) d\tau$, nous utilisons le formalisme (5.22) appliqué au cas où $V(y) =$

$\delta(y - r)$. Dans ce cas, le hamiltonien est $H_\lambda = -\frac{1}{2}\frac{d^2}{dy^2} + \lambda\delta(y - r)$, les deux solutions indépendantes $u_s(y)$ et $v_s(y)$ recherchées vérifient donc

$$-\frac{1}{2}\psi''(y) + s\psi(y) = 0, \text{ avec } y \in [0, r[\cup]r, +\infty[, \quad (5.27)$$

avec les conditions aux limites

$$\begin{cases} u_s(0) = 0 \\ u_s(r^+) = u_s(r^-) \\ \frac{1}{2}(u'_s(r^+) - u'_s(r^-)) = \lambda u_s(r), \end{cases} \quad \begin{cases} \lim_{y \rightarrow \infty} v_s(y) = 0 \\ v_s(r^+) = v_s(r^-) \\ \frac{1}{2}(v'_s(r^+) - v'_s(r^-)) = \lambda v_s(r). \end{cases} \quad (5.28)$$

Les conditions aux limites imposées en r proviennent du pic δ présent dans le hamiltonien. Il est possible de résoudre par morceaux l'équation différentielle (5.27) avec les conditions (5.28), on trouve alors

$$u_s(y) = \begin{cases} A \sinh(\sqrt{2s}y) & y \leq r \\ A \left(\sinh(\sqrt{2s}y) + \frac{2\lambda \sinh(\sqrt{2s}r)}{\sqrt{2s}} \sinh(\sqrt{2s}(y - r)) \right) & y > r \end{cases} \quad (5.29)$$

$$v_s(y) = \begin{cases} B \left(e^{-\sqrt{2s}(y-r)} + \frac{2\lambda}{\sqrt{2s}} \sinh(\sqrt{2s}(r - y)) \right) & y \leq r \\ B e^{-\sqrt{2s}(y-r)} & y > r, \end{cases} \quad (5.30)$$

où les constantes de normalisations A et B peuvent être calculées mais ne seront pas utiles pour la suite. Le wronskien est alors donné par

$$W = u'_s(y)v_s(y) - u_s(y)v'_s(y) = ABe^{\sqrt{2s}r} \left[(1 - e^{-2\sqrt{2s}r})\lambda + \sqrt{2s} \right]. \quad (5.31)$$

La combinaison de ces différents résultats (5.29), (5.30) et (5.31), (5.22) donne directement la double transformée de Laplace de la densité à partir du maximum

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \rho_{\max}(r,t)} \rangle = \frac{1}{s} \left(\frac{\sqrt{2s} + \lambda(1 - e^{-\sqrt{2s}r})^2}{\sqrt{2s} + \lambda(1 - e^{-2\sqrt{2s}r})} \right)^2. \quad (5.32)$$

À partir de ce résultat (5.32), il est possible d'obtenir les différents moments $\langle \rho_{\max}^k(r, t) \rangle$, avec k entier en faisant un développement en puissance de λ .

$$\begin{aligned} \int_0^\infty dt e^{-st} \langle e^{-\lambda \rho_{\max}(r,t)} \rangle = \\ 1 + \sum_{k=1}^\infty (-\lambda)^k \frac{8e^{-\sqrt{2s}r}}{(2s)^{k/2+1}} (1 - e^{-\sqrt{2s}r})^2 (1 + ke^{-\sqrt{2s}r})(1 - e^{-2\sqrt{2s}r})^{k-2}. \end{aligned} \quad (5.33)$$

On lit directement dans le développement (5.33) les transformées de Laplace par rapport à t des différents moments $\langle \rho_{\max}^k(r, t) \rangle$

$$\begin{aligned} \int_0^\infty dt \langle \rho_{\max}^k(r, t) \rangle e^{-st} = \\ \frac{8k!}{(2s)^{k/2+1}} \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} \left((2l-1)e^{-(2l+1)\sqrt{2s}r} + (k-2(l+1))e^{-2(l+1)\sqrt{2s}r} \right). \end{aligned} \quad (5.34)$$

Il est alors possible d'inverser ces transformées de Laplace qui font apparaître les fonctions $\Phi^{(j)}(x)$ introduit dans l'appendice A

— Moments de la densité à partir du maximum pour le brownien libre —

$$\begin{aligned} \langle \rho_{\max}^k(r, 1) \rangle = & \quad (5.35) \\ 8k! \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} [(2l+1)\Phi^{(k+1)}((2l+1)r) + (k-2(l+1))\Phi^{(k+1)}(2(l+1)r)] . \end{aligned}$$

Pour $k = 1$, on retrouve bien le résultat obtenu par la méthode des propagateurs (3.18). On définit $P_t(\rho, r)$ la densité de probabilité de la densité à partir du maximum $\rho_{\max}(r, t)$ (2.17) c'est-à-dire $P_t(\rho, r)d\rho$ est la probabilité que $\rho_{\max}(r, t)$ soit compris entre ρ et $\rho + d\rho$. En inversant la transformée de Laplace par rapport à λ de l'équation (5.32), on obtient la transformée de Laplace par rapport à t de $P_t(\rho, r)$

— Loi de la densité à partir du maximum pour le brownien libre —

$$\begin{aligned} \int_0^\infty e^{-st} P_t(\rho, r) dt = & \delta(\rho) \frac{(e^{-\sqrt{2s}r} - 1)^2}{s(1 + e^{-\sqrt{2s}r})^2} \quad (5.36) \\ & + \frac{e^{-\frac{\rho\sqrt{2s}e^{\sqrt{2s}r}}{2\sinh(\sqrt{2s}r)}}}{\cosh^3\left(\frac{r\sqrt{2s}}{2}\right)} \left(\frac{e^{\frac{r\sqrt{2s}}{2}}}{\sqrt{2s}} + \frac{\rho e^{\sqrt{2s}r}}{4\sinh(r\sqrt{2s})\sinh\left(\frac{r\sqrt{2s}}{2}\right)} \right) . \end{aligned}$$

La densité de probabilité est composée d'un poids concentré en $\rho = 0$ avec le pic $\delta(\rho)$ en plus d'une densité continue pour $\rho > 0$. On peut donc ainsi écrire

$$P_t(\rho, r) = F_W(r, t)\delta(\rho) + p_t(\rho, r) , \quad (5.37)$$

où on lit directement dans (5.36) les transformées de Laplace respective des deux termes. L'inversion de Laplace du premier terme donne

$$F_W(r, t) = 1 + \sum_{l=1}^{\infty} 4l(-1)^l \operatorname{erfc}(lr/\sqrt{2t}) . \quad (5.38)$$

On reconnaît dans cette formule la loi de la largeur pour un brownien libre [56] c'est-à-dire la probabilité que la largeur $W(t) = \max_{\tau \in [0, t]} x(\tau) - \min_{\tau \in [0, t]} x(\tau)$ soit plus petite que r . En effet, si cette largeur est plus petite que r , le temps passé par le brownien à une distance r du maximum est bien nul ce qui conduit au pic delta dans la formule (5.36). Le second terme de (5.37), dont la dépendance en temps est triviale via l'autosimilarité du brownien $p_t(\rho, r) = t^{-\frac{1}{2}} p_1(\rho/\sqrt{t}, r/\sqrt{t})$, est une fonction continue de ρ dont la transformée de Laplace par rapport à t est le second terme de l'équation (5.36). Dans la figure 5.2, nous avons comparé les résultats numériques pour trois valeurs de r (les points) ainsi que nos résultats analytiques (lignes) en inversant la transformée de Laplace (5.36) numériquement et observons une très bonne concordance. Il est possible à partir de la transformée de Laplace (5.36) de remonter aux comportements asymptotiques de $p_1(\rho, r)$

$$p_1(\rho, r) = \begin{cases} \frac{\rho}{4r^2} e^{-\frac{\rho}{2r}} & \text{pour } r \rightarrow 0 \\ (1 - F_W(r, 1)) r e^{-\rho r} & \text{pour } r \rightarrow \infty . \end{cases} \quad (5.39)$$

$$(5.40)$$

En particulier, le comportement à r petit est bien cohérent avec les moments (5.35) à r petit

$$\langle \rho_{\max}^k(r, 1) \rangle = \int_0^\infty d\rho \rho^k P_1(\rho, r) \underset{r \rightarrow 0}{\sim} 2^k (k+1)! r^k . \quad (5.41)$$

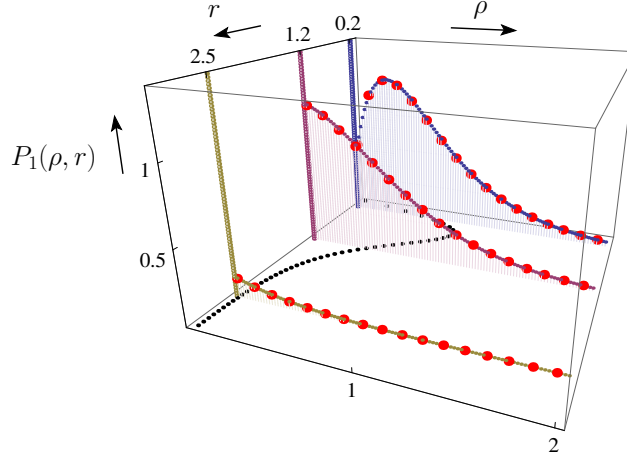


FIGURE 5.2: Tracé de $P_1(\rho, r)$ comme une fonction de ρ pour trois valeurs de r (en vert $r = 2.5$, en violet $r = 1.2$ et en bleu $r = 0.2$). Les lignes à $\rho = 0$ sont les pics $\delta(\rho)$ de l'équation (5.36). Les lignes correspondent aux résultats analytiques obtenus après une inversion de Laplace numérique du second terme de (5.36). Les points rouges correspondent aux résultats numériques en moyennant sur 10^7 marches aléatoires de type (2.2). Nous avons ajouté dans le plan $z = 0$ en pointillés noirs la valeur moyenne $\langle \rho_{\max}(r, t) \rangle = \int_0^\infty d\rho \rho P_t(\rho, r)$ (3.18).

Remarquons également que le terme $1 - F_W(r, 1)$ dans le développement à grand r représente la probabilité que la largeur du brownien soit plus grande que r ce qui est cohérent avec (5.37). Il est également possible d'obtenir et d'interpréter la fonction limite de $p_1(\rho, r)$ à ρ petit. Pour cela, prenons la limite $\rho \rightarrow 0$ dans la transformée de Laplace de $p_1(\rho, r)$ que l'on peut lire dans (5.36) (pour $r > 0$)

$$\int_0^\infty e^{-st} p_t(0, r) dt = \frac{8e^{-\sqrt{2s}r}}{(1 + e^{-\sqrt{2s}r})^3 \sqrt{2s}} = \frac{4}{\sqrt{2s}} \sum_{l=1}^\infty l(l+1)(-1)^{l+1} e^{-\sqrt{2s}lr}. \quad (5.42)$$

Il est alors possible d'inverser cette transformée de Laplace

$$p_t(0, r) = 2\sqrt{\frac{2}{\pi t}} \sum_{l=1}^\infty (-1)^{l+1} l(l+1) e^{-\frac{l^2 r^2}{2t}}, \quad r > 0. \quad (5.43)$$

Nous décomposons alors, pour l'interprétation, cette somme en deux parties distinctes

$$p_t(0, r) = \frac{1}{2} \partial_r F_W(r, t) + p_J(r, t), \quad r > 0. \quad (5.44)$$

où la première partie est

$$\frac{1}{2} \partial_r F_W(r, t) = \frac{1}{2} \partial_r \left[1 + \sum_{l=1}^\infty 4l(-1)^l \operatorname{erfc}(lr/\sqrt{2t}) \right] = 2\sqrt{\frac{2}{\pi t}} \sum_{l=1}^\infty (-1)^{l+1} l^2 e^{-\frac{l^2 r^2}{2t}}, \quad (5.45)$$

avec $F_W(r, t)$ la loi de probabilité de la largeur d'un brownien libre (5.38). Le second terme $p_J(r, t)$ est alors

$$p_J(r, t) = 2\sqrt{\frac{2}{\pi t}} \sum_{l=1}^\infty (-1)^{l+1} l e^{-\frac{l^2 r^2}{2t}}. \quad (5.46)$$

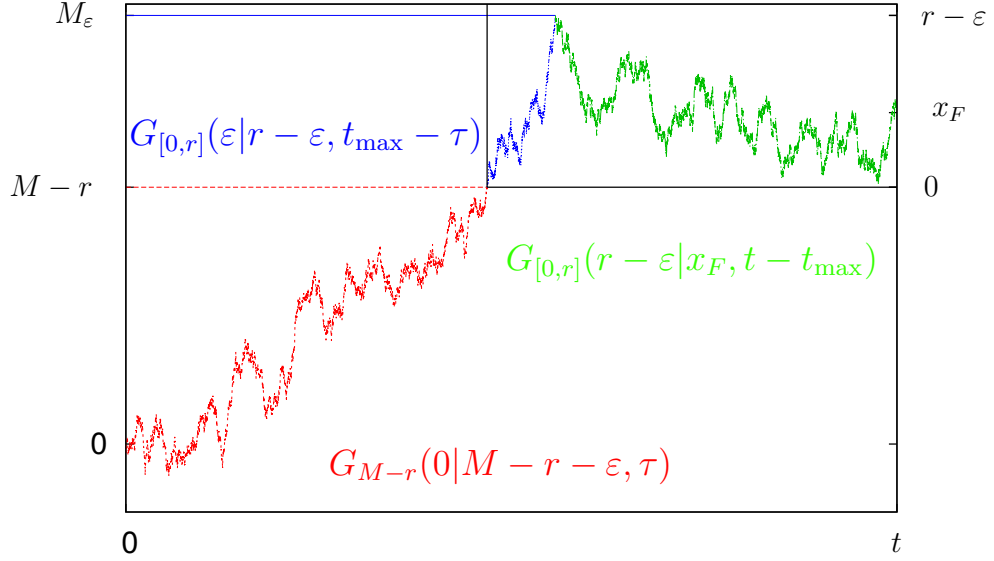


FIGURE 5.3: Les chemins qui contribuent à $p_J(r, t)$ sont les chemins (ainsi que leurs symétriques) qui passent en $M - r$ au temps τ pour la première fois, puis reste toujours plus grand que $M - r$ après τ avec M le maximum du brownien. Remarquons que la trajectoire après le temps τ peut être vue comme un méandre de maximum r .

Dans le domaine de Laplace, cette somme peut être réécrite

$$\int_0^\infty dt p_J(r, t) e^{-st} = \frac{4}{\sqrt{2s}} \sum_{l=1}^\infty (-1)^{l+1} l e^{-\sqrt{2s}rl} = \frac{1}{\sqrt{2s} \cosh^2\left(\frac{r\sqrt{2s}}{2}\right)}. \quad (5.47)$$

Il est possible de montrer que cette somme est égale au produit de propagateurs (définis en 3.1)

$$\frac{1}{Z(\varepsilon)} \int_r^\infty dM \tilde{G}_{M-r}(0|M-r-\varepsilon, s) \tilde{G}_{[0,r]}(\varepsilon|r-\varepsilon, s) \int_0^r dx_F \tilde{G}_{[0,r]}(r-\varepsilon|x_F, s), \quad (5.48)$$

avec la constante de normalisation $Z(\varepsilon) = 4\varepsilon^2$ (car $\int_0^\infty dr p_J(r, t) = 1$). On lit alors dans le produit (5.48) le type de chemin représentée par la contribution $p_J(r, t)$. Il s'agit des chemins qui passe une et une seule fois le point situé à une distance r du maximum (comme la trajectoire de la figure 5.3). Remarquons que $p_J(r, t)$ peut se réécrire comme

$$p_J(r, t) = \int_0^t d\tau \frac{1}{\sqrt{\pi\tau(t-\tau)}} p_{\max}(r, t-\tau), \quad (5.49)$$

avec $p_{\max}(r, t)$ la distribution de probabilité du maximum du méandre

$$p_{\max}(r, t) = \frac{d}{dr} \left(\frac{2\sqrt{2\pi t}}{r} \sum_{n=0}^\infty e^{-\frac{\pi^2}{2r^2}(2n+1)^2 t} \right). \quad (5.50)$$

Pour comprendre la formule (5.49), il suffit de transformer la trajectoire de la figure (5.3) en symétrisant la partie postérieure au temps τ par rapport à la droite $y_M - r$. Après cette transformation, τ devient le temps auquel la nouvelle trajectoire, qui est un brownien

libre, a atteint son maximum. Ce temps est donc donné par la loi de l'arc sinus (5.1) que l'on retrouve donc dans (5.49). Enfin, il n'y a aucune contrainte sur la trajectoire avant le temps τ tandis que la trajectoire sur $[\tau, t]$ doit être un méandre de maximum r ce qui explique le terme $p_{\max}(r, t)$ dans (5.49). Cette analyse n'est valable que pour $r > 0$. Dans le cas $r = 0$, on s'attend à une contribution additionnelle qui provient des cas où le maximum est unique. Ainsi,

$$p_t(0, r) = \alpha \delta(r) + \frac{1}{2} \partial_r F_W(r, t) + p_J(r, t) . \quad (5.51)$$

Remarquons que le terme ajouté $t^{-1/2} \delta(rt^{-1/2}) = \delta(r)$ ne contredit pas la forme d'échelle du brownien. Pour obtenir le paramètre α , nous utilisons le comportement asymptotique tiré de (5.36) dans la limite où $\rho \rightarrow 0$, $r \rightarrow 0$ avec ρ/r fixé

$$p_t(\rho, r) \underset{\substack{\rho \rightarrow 0, r \rightarrow 0 \\ \rho/r = \text{cte}}}{\sim} \frac{\rho}{4r^2} e^{-\frac{\rho}{2r}} . \quad (5.52)$$

Cette forme d'échelle montre que quand ρ tend vers 0, $p_t(\rho, r)$ est de plus en plus piquée autour de $r = 0$ ce qui conduit au pic $\delta(r)$ quand $\rho = 0$ (5.51). Puisque

$$\int_0^\infty dr \frac{\rho}{4r^2} e^{-\frac{\rho}{2r}} = \frac{1}{2} , \quad (5.53)$$

on en déduit que $\alpha = 1/2$. Finalement, on trouve donc

$$p_t(0, r) = \frac{1}{2} \delta(r) + \frac{1}{2} \partial_r F_W(r, t) + p_J(r, t) . \quad (5.54)$$

Même s'il est difficile d'interpréter cette relation sur le brownien à cause des problèmes de régularisations déjà mentionnés plusieurs fois, il est possible de discuter précisément la formule (5.54) au moins pour la marche aléatoire discrète en temps et en espace

$$\begin{cases} X_0 = 0 \\ X_i = X_{i-1} + \eta_i, i \in [1, N] \end{cases} \quad (5.55)$$

où les η_i sont des variables $\{-1, +1\}$ chacun avec probabilité 1/2. Dans la limite où N tend vers l'infini, $X_{\lfloor \tau N \rfloor} / \sqrt{N}$ tend vers un brownien libre $\tau \in [0, 1]$. Dans ce cas, $p_t(0, r)$ représente la probabilité qu'il y ait un seul point à distance r du maximum. Plus précisément, dans la limite de grand n , la probabilité $P_{k,N}(m)$ que le marcheur discret après N sauts ± 1 ait visité le site m exactement k fois est

$$P_{k,N}(m) \underset{n \rightarrow \infty}{=} \frac{1}{\sqrt{N}} p_1 \left(\frac{k}{\sqrt{N}}, \frac{m}{\sqrt{N}} \right) . \quad (5.56)$$

Il n'y a qu'un seul point à une distance $r = 0$ du maximum évidemment que si celui-ci est unique ce qui arrive avec une probabilité, qui est un résultat de notre calcul, 1/2 et donne le terme $1/2 \delta(r)$. Enfin, il n'y a qu'un seul point à une distance $r > 0$ du maximum dans seulement deux cas :

- si le minimum est à une distance r du maximum c'est-à-dire la largeur $W(t)$ du brownien est r et que ce minimum est unique ce qui arrive également avec une probabilité 1/2. Cela aboutit au terme $\frac{1}{2} \partial_r F_W(r, t)$
- si la trajectoire, qui a une largeur supérieure à r , n'est passé qu'une unique fois à la position r du maximum comme pour les trajectoires du type de la figure 5.3 dans la limite continue. Ce cas aboutit ainsi au terme $p_J(r, t)$.

Cas du pont

Dans le cas du pont, en utilisant la construction de Vervaat schématisée dans la figure 3.2, nous avons montré l'égalité en loi entre $\rho_{\max,B}(r,t)$ et le temps local de l'excursion $T_{loc,E}(r,t)$ (3.20). En utilisant le résultat général (5.26) appliqué à la densité à partir du maximum c'est-à-dire avec le potentiel $V(x) = \delta(x - r)$

$$\langle e^{-\lambda \rho_{\max,B}(r,t)} \rangle = \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{2\pi} t^{3/2}}{2\varepsilon^2} \langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle, \quad (5.57)$$

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dx^2} + \lambda \delta(x - r) + V_{\text{wall}}(x), \quad (5.58)$$

où $V_{\text{wall}}(x)$ est une barrière de potentiel qui confine la particule sur le demi-axe positif : $V_{\text{wall}}(x) = 0$ pour $x \geq 0$ et $V_{\text{wall}}(x) = +\infty$ pour $x < 0$. Il est alors possible de calculer le numérateur de (5.57) en utilisant les fonctions propres $\psi_k(x)$ d'énergie $E_k = \frac{k^2}{2}$ du hamiltonien H_λ données par

$$\psi_k(x) = \begin{cases} 0 & x \leq 0 \\ a(k, \lambda) \sin(kx) & 0 \leq x \leq r \\ a(k, \lambda) \left(\sin(kx) + \frac{2\lambda}{k} \sin(kr) \sin(k(x-r)) \right) & r \leq x, \end{cases} \quad (5.59)$$

où les amplitudes $a(k, \lambda)$ assurent la condition de normalisation

$$|a(k, \lambda)|^{-2} = \frac{\pi}{2} \left[1 + \frac{4\lambda}{k} \sin(kr) \left(\cos(kr) + \frac{\lambda}{k} \sin(kr) \right) \right]. \quad (5.60)$$

Avec ces résultats, on obtient directement

— Transformée de Laplace de la densité à partir du maximum pour le pont —

$$\begin{aligned} \langle e^{-\lambda \rho_{\max,B}(r,t)} \rangle &= \langle e^{-\lambda T_{loc,E}(r,t)} \rangle \\ &= \sqrt{\frac{2t^3}{\pi}} \int_0^\infty \frac{k^2 e^{-\frac{k^2 t}{2}}}{1 + \frac{4\lambda}{k} \sin(kr) \left(\frac{\lambda}{k} \sin(kr) + \cos(kr) \right)} dk. \end{aligned} \quad (5.61)$$

Remarquons que cette expression est un résultat nouveau pour le temps local de l'excursion. De cette expression, il est possible d'obtenir les différents moments de la densité à partir du maximum $\langle \rho_{\max,B}^k(r,t) \rangle$ (pour $k \in \mathbb{N}$) via un développement en puissance de λ fait comme le montre l'appendice D. On retrouve bien dans ce cas les résultats obtenus avec la méthode des propagateurs (3.25). Il est également possible d'obtenir les différentes transformées de Laplace

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \rho_{\max,B}(r,t)} \rangle = \frac{1}{s} - \frac{2\sqrt{2\pi}\sqrt{2s}}{e^{2\sqrt{2s}r} \left(1 + \frac{\sqrt{2s}}{\lambda} \right) - 1}, \quad (5.62)$$

— Loi de la densité à partir du maximum pour le pont —

$$\int_0^\infty dt e^{-st} P_t(\rho, r) = \delta(\rho) \left(\frac{1}{s} - \frac{4\sqrt{\pi s}}{e^{2\sqrt{2s}r} - 1} \right) + \frac{\sqrt{2\pi s}}{\sinh^2(\sqrt{2s}r)} \exp\left(-\frac{\sqrt{2s}\rho e^{\sqrt{2s}r}}{2 \sinh(\sqrt{2s}r)} \right). \quad (5.63)$$

Le premier terme est la transformée de Laplace de la loi de la largeur d'un pont $F_W^B(r, t)$ [30]

$$F_W^B(r, t) = 1 + 2 \sum_{l=1}^{\infty} e^{-\frac{2l^2 r^2}{t}} \left(1 - \frac{4l^2 r^2}{t} \right), \quad (5.64)$$

tandis que le second terme, qu'on peut trouver dans [82] et [75], est la transformée de Laplace de $p_t^B(\rho, r)$ qui est la partie continue de la densité de probabilité de $\rho_{\max, B}(r, t)$. Ainsi, comme précédemment (5.37), la densité de probabilité de $\rho_{\max, B}(r, t)$ s'écrit

$$P_t^B(\rho, r) = F_W^B(r, t) \delta(\rho) + p_t^B(\rho, r), \quad (5.65)$$

Il est possible de montrer [149, 150] que les moments (3.25) permettent de remonter à la distribution

$$p_t^B(\rho, r) = 1 - 2 \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \binom{j-1}{k} e^{-(\rho+2rj)^2/2} (-\rho)^k H_{k+2}(\rho+2rj)/k!, \quad (5.66)$$

où les H_n sont les polynômes d'Hermite (20). Dans le cas du pont, les comportements asymptotiques se réécrivent

$$p_1^B(\rho, r) = \begin{cases} \frac{\rho}{4r^2} e^{-\frac{\rho}{2r}} & r \rightarrow 0 \\ (1 - F_W^B(r, 1)) 2r e^{-2\rho r} & r \rightarrow \infty \end{cases} \quad (5.67)$$

Le comportement à r petit est obtenu par exemple en étudiant la limite de grand λ dans l'équation (5.61). On retrouve la formule du brownien libre ce qui montre encore une fois que le comportement proche du maximum ne dépend pas des conditions aux bords. Ce comportement est également bien cohérent avec les comportements à r petit des moments (3.25). De la même façon que pour le brownien libre, le terme $1 - F_W^B(r, 1)$ dans le développement à grand r représente la probabilité que la largeur du pont soit plus grande que r ce qui est cohérent avec (5.65). Il est également possible d'obtenir et d'interpréter la fonction limite de $p_1^B(\rho, r)$ à ρ petit. Pour cela, prenons la limite $\rho \rightarrow 0$ dans la transformée de Laplace de $p_1^B(\rho, r)$ que l'on peut lire dans (5.63) (pour $r > 0$), la contribution $p_J(r, t)$ ne pouvant plus exister à cause des conditions aux limites périodiques qui empêche les trajectoires du type de la figure 5.3 d'exister dans ce cas

$$p_t^B(0, r) = \frac{1}{2} \delta(r) + \frac{1}{2} \partial_r F_W^B(r, t). \quad (5.68)$$

5.2.2 Applications aux cas où $V(x) = x^\alpha$

Nous allons désormais étudier la famille de fonctionnelles $T_\alpha(t)$ pour laquelle $V(x) = x^\alpha$ avec $\alpha \in]2, \infty[$

$$T_\alpha(t) = \int_0^t d\tau (x_{\max} - x(\tau))^\alpha, \quad (5.69)$$

où x_{\max} est le maximum du brownien libre $x(\tau)$ pour $\tau \in [0, t]$. Ces fonctionnelles ont plusieurs applications que nous discuterons par la suite. On peut déduire rapidement la dépendance temporelle de ces fonctionnelles $T_\alpha(t) = t^{1+\alpha/2} T_\alpha(t=1)$ à partir de l'auto-similarité du brownien. Il est également possible de considérer ces fonctionnelles dans le cas du pont

$$T_{\alpha, B}(t) = \int_0^t d\tau (x_{\max, B} - x_B(\tau))^\alpha, \quad (5.70)$$

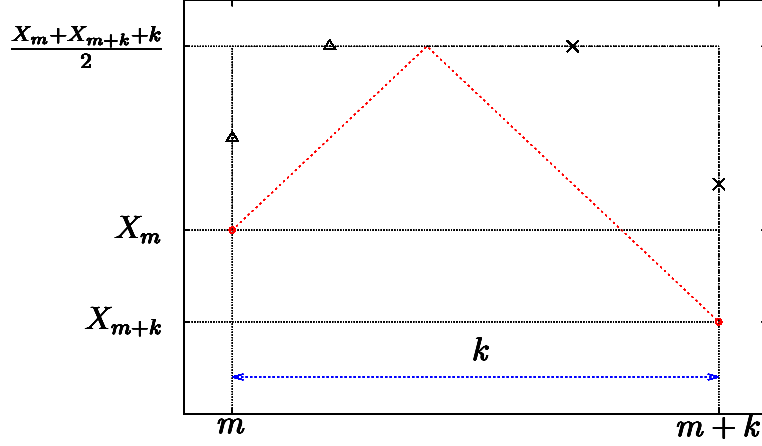


FIGURE 5.4: En révélant les positions X_k et X_{m+k} , on sait alors que le maximum entre ces deux positions ne peut excéder $(X_k + X_{m+k} + k)/2$.

où $x_{\max,B}$ est le maximum du pont $x_B(\tau)$ pour $\tau \in [0, t]$ et dont la dépendance temporelle est également $T_{\alpha,B}(t) = t^{1+\alpha/2}T_{\alpha,B}(t=1)$. Cette famille de fonctionnelles extrapole en particulier trois problèmes :

— pour $\alpha = -1$:

de telles fonctionnelles apparaissent dans l'algorithme optimal de recherche du maximum pour une marche aléatoire [119]. Considérons la marche aléatoire que nous avons déjà introduit en (5.55)

$$\begin{cases} X_0 = 0 \\ X_i = X_{i-1} + \eta_i, i \in [1, n] \end{cases} \quad (5.71)$$

où les η_i sont des variables aléatoires $\{-1, +1\}$ chacun avec probabilité $1/2$. Rappelons que dans la limite des grands n , $X_{\lfloor \tau n \rfloor} / \sqrt{n}$ tend vers un brownien libre avec $\tau \in [0, 1]$. Nous recherchons alors la méthode optimale pour trouver le maximum de cette marche aléatoire $M_n = \max\{X_i, 0 \leq i \leq n\}$ en minimisant le nombre de positions X_k révélées pour trouver ce maximum. Nous nous autorisons à faire autant de calculs que nécessaires avec les positions X_k examinées, seul compte le nombre de positions révélées dont nous avons eu besoin pour trouver le maximum. On note A_n l'ensemble des algorithmes permettant de trouver le maximum. Pour un algorithme a de cet ensemble, on note $C(a)$ le coût de cet algorithme pour trouver la maximum d'une configuration c'est-à-dire le nombre de positions X_k que l'algorithme a dû révéler pour trouver le maximum avec certitude. Le premier algorithme auquel on peut penser et de révéler toutes les positions X_k pour $1 \leq k \leq n$ dont il est facile de tirer le maximum. Le coût d'un tel algorithme est alors n . Il est en réalité possible de faire mieux en utilisant la corrélation qui existe entre les différentes positions du marcheur aléatoire comme le montre la figure 5.4. On donne par exemple en figure 5.5 un exemple de méthode, que l'on nomme algorithme u , pour trouver sur cet exemple de trajectoire de 14 pas le maximum $M_{14} = 7$ en révélant seulement 4 positions. Bien sûr, certaines trajectoires nécessitent d'examiner plus ou moins de position pour trouver le maximum. Par exemple, la marche aléatoire constituée uniquement de saut $+1$ ne peut nécessiter que l'observation de $X_n = n$ pour trouver son maximum tandis que la marche constituée alternativement de saut ± 1 ($X_{2k} = 0$ et $X_{2k+1} = 1$) nécessite de révéler au moins $\lfloor \frac{n}{2} \rfloor + 1$ positions pour être sûr qu'aucune position impaire ne soit 2. Ainsi, le coût $C(a)$ d'un algorithme $a \in A_n$ est une variable aléatoire qui varie d'une trajectoire à l'autre, on note sa

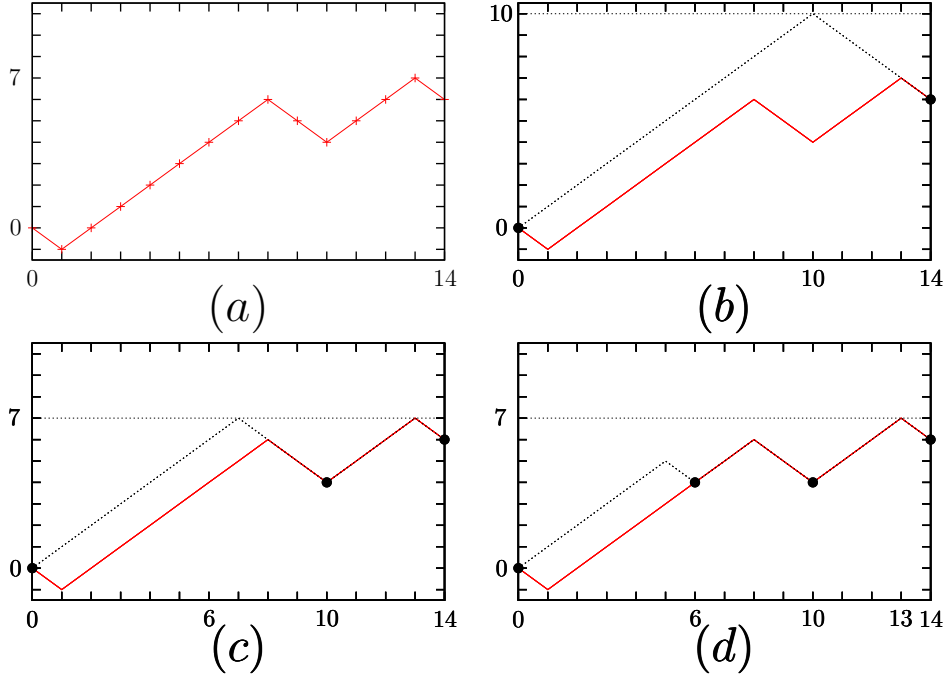


FIGURE 5.5: Exemple de l'algorithme u pour trouver $M_{14} = 7$ en révélant seulement 4 positions. **(a)** Exemple de trajectoire dont on cherche le maximum. Initialement, on sait seulement que cette trajectoire démarre en $X(0) = 0$. Puisque celle-ci est composée de saut ± 1 , on déduit que le maximum est compris entre $0 \leq M_{14} \leq 14$, et si $M_{14} = 14$ la position du maximum serait 14 (la marche aléatoire serait alors composée que de sauts $+1$). On examine alors la position 14. **(b)** La position révélée est $X_{14} = 6$. On en déduit donc (cf. figure 5.4) que $6 \leq M_{14} \leq 10$ et si $M_{14} = 10$ alors la position du maximum serait 10 (trajectoire en ligne pointillée). On examine alors la position 10. **(c)** La seconde position révélée est $X_{10} = 4$. On en déduit donc $6 \leq M_{14} \leq 7$ et si $M_{14} = 7$ alors la position du maximum serait 6 ou 13 (trajectoire en ligne pointillée). On examine alors la position 6. **(d)** La troisième position révélée est $X_6 = 4$. On en déduit donc $6 \leq M_{14} \leq 7$ et si $M_{14} = 7$ alors la position du maximum serait 13 (trajectoire en ligne pointillée). On examine alors la position 13. On trouve finalement le maximum $M_{14} = X_{13} = 7$ en ayant révélé seulement 4 positions.

moyenne sur les différentes trajectoires possibles $\langle C(a) \rangle$. Odlyzko [119] a étudié le coût moyen minimal pour ces algorithmes et a montré que dans la limite des n grands, ce coût moyen est

$$\min_{a \in A_n} \langle C(a) \rangle = c_0 \sqrt{n} + o(\sqrt{n}) , \quad (5.72)$$

avec c_0 une constante donnée par

$$c_0 = \langle I \rangle , \quad I = \frac{1}{2} \int_0^1 \frac{d\tau}{x_{\max} - x(\tau)} = \frac{1}{2} T_{\alpha=-1}(t=1) . \quad (5.73)$$

On trouvera plus d'indications en appendice E. Odlyzko a tout d'abord trouvé une expression de c_0 sous la forme d'une double intégrale qui a par la suite été calculée indépendamment par Hwang [85] et Chassaing [26]

$$c_0 = \sqrt{\frac{8}{\pi}} \log 2 = 1.1061 \dots . \quad (5.74)$$

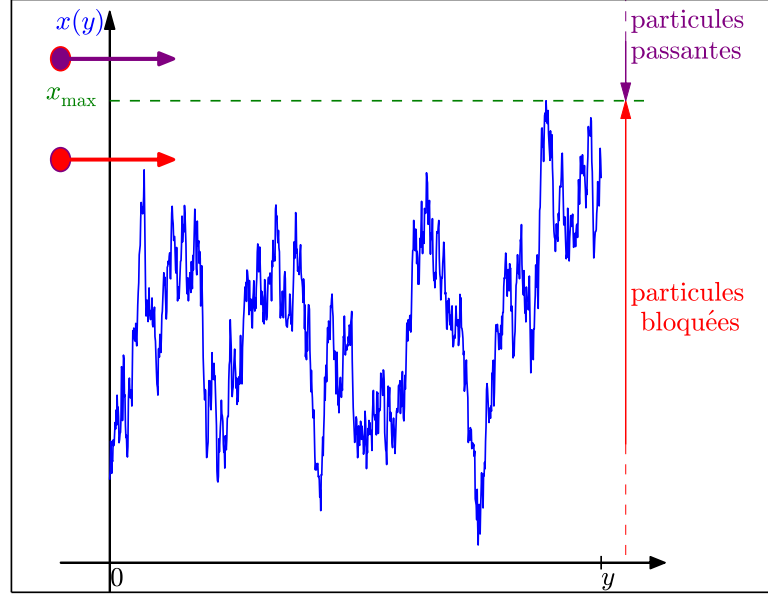


FIGURE 5.6: Particules balistiques traversant un potentiel aléatoire brownien.

Par la suite, les travaux de Chassaing *et al.* [28] ont généralisé la limite (5.72) à la loi complète

$$\lim_{n \rightarrow \infty} \left(\min_{a \in A_n} \Pr . \left[\frac{C(a)}{\sqrt{n}} \geq x \right] \right) = \Pr . [I \geq x] , \quad (5.75)$$

où I est la variable aléatoire (5.73). La distribution ainsi que les différents moments de I ont alors été étudié [28] en utilisant des méthodes issues des probabilités. Nous allons retrouver ces résultats à partir de notre méthode issue de l'intégrale de chemin.

— pour $\alpha = -1/2$:

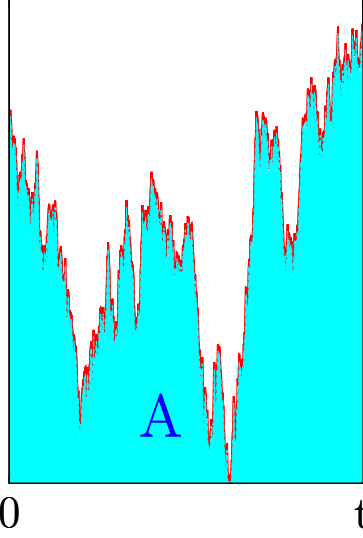
Une autre application intéressante est le cas où $V(x) = 1/\sqrt{x}$. Ce cas décrit par exemple le temps maximal pour une particule balistique (de masse unité) pour traverser un potentiel aléatoire brownien comme sur la figure 5.6. En effet, imaginons une particule classique dans un potentiel brownien $x(y)$ sur $y \in [0, 1]$ (la variable d'espace y joue le rôle du temps du brownien comme nous l'avons défini en 2.1.2). Cette particule de masse unité est déposée avec une énergie E au point $y = 0$. Par conservation de l'énergie

$$\frac{1}{2} \left(\frac{dy}{dt} \right)^2 + x(y) = E . \quad (5.76)$$

La particule classique peut traverser la région $y \in [0, 1]$ seulement si son énergie est plus grande que la plus haute barrière de potentiel sur ce segment c'est-à-dire $E > x_{\max} = \max_{y \in [0, 1]} x(y)$. Le temps nécessaire pour une telle particule pour traverser cette région est alors donné par

$$\mathcal{T}(E) = \frac{1}{\sqrt{2}} \int_0^1 \frac{dy}{\sqrt{E - x(y)}} . \quad (5.77)$$

aire sous le double méandre



aire sous l'excursion

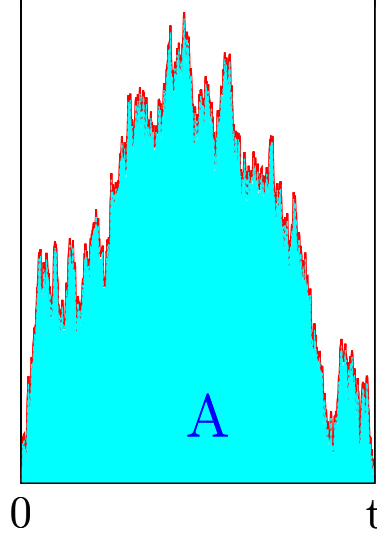


FIGURE 5.7: (gauche) La fonctionnelle $T_{\alpha=1}(t)$ entre le brownien libre et son maximum c'est-à-dire l'aire A sous le double méandre. (droite) La fonctionnelle $T_{\alpha=1,B}(t)$ mesure la surface entre le pont et son maximum c'est-à-dire l'aire A sous l'excursion.

Si l'on envoie un faisceau de particules possédant des énergies différentes, seules les particules d'énergie $E > x_{\max}$ traversent la région $y \in [0, 1]$. La particule la plus lente pour traverser est celle d'énergie $E = x_{\max}$, cela nécessitera un temps

$$\mathcal{T}_{\max} = \max_{E \geq x_{\max}} \mathcal{T}(E) = \mathcal{T}(x_{\max}) = \frac{1}{\sqrt{2}} \int_0^1 \frac{dy}{\sqrt{x_{\max} - x(y)}} \quad (5.78)$$

qui est donné par la fonctionnelle du maximum $\frac{1}{\sqrt{2}}T_{\alpha=-1/2}(t=1)$.

— pour $\alpha = 1$:

dans ce cas, la fonctionnelle $T_{\alpha=1}(t)$ mesure la surface sur $[0, t]$ comprise entre le brownien et la droite $y = x_{\max}$ c'est-à-dire, en utilisant la transformation expliquée en figure 5.1, l'aire sous le double méandre [88, 106]. Dans le cas du pont, grâce à la transformation de Vervaat [161] expliquée en figure 3.2, $T_{\alpha=1,B}(t)$ mesure la surface sous une excursion. Ces deux aires sont représentées sur la figure 5.7. Cette aire sous l'excursion est distribuée pour $t = 1$ suivant la loi d'Airy [105, 106, 148, 151].

$$f(A) = \frac{2\sqrt{6}}{A^{10/3}} \sum_{k=1}^{\infty} e^{-b_k/A^2} b_k^{2/3} U(-5/6, 4/3, b_k/A^2), \quad (5.79)$$

où $b_k = 2\alpha_k^3/27$ avec α_k la k -ième amplitude des zéros de la fonction d'Airy (12) sur l'axe négatif ($\alpha_1 = 2.3381\dots$, $\alpha_2 = 4.0879\dots$, $\alpha_3 = 5.5205\dots$) et $U(a, b, z)$ la fonction hypergéométrique confluyente de Tricomi (18). Cette distribution se retrouve dans différents domaines comme l'informatique théorique [34, 62, 100] ou encore la statistique d'extrême pour des interfaces élastiques [103, 105, 106].

Dans un premier temps, il est facile à partir de la relation (2.19) de calculer la moyenne de la famille de fonctionnelles pour les différents mouvements browniens. On trouve alors dans le cas du brownien libre avec $\alpha \in]2, +\infty[$

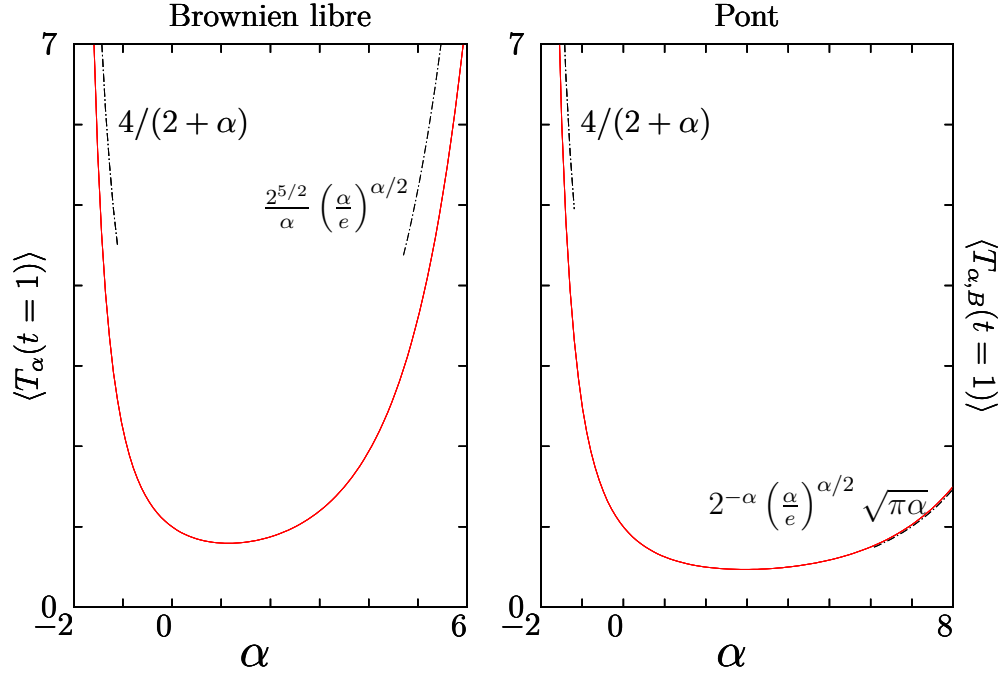


FIGURE 5.8: Tracé de $\langle T_\alpha(t=1) \rangle$ (5.80) et $\langle T_{\alpha,B}(t=1) \rangle$ (5.81) en fonction de $\alpha > -2$. $\langle T_\alpha(t=1) \rangle$ a un comportement intéressant en fonction de α , elle diverge pour $\alpha \rightarrow -2$ comme $\langle T_\alpha(1) \rangle \sim 4/(2+\alpha)$, décroît pour atteindre un minimum pour $\alpha \approx 1.148$ puis croît jusqu'à être équivalente à $\langle T_\alpha(1) \rangle \sim \frac{2^{5/2}}{\alpha} \left(\frac{\alpha}{e}\right)^{\alpha/2}$ quand $\alpha \rightarrow \infty$. De la même façon, $\langle T_{\alpha,B}(t=1) \rangle$ diverge pour $\alpha \rightarrow -2$ comme $\langle T_{\alpha,B}(1) \rangle \sim 4/(2+\alpha)$, décroît pour atteindre un minimum pour $\alpha \approx 2.960$ puis croît jusqu'à être équivalente à $\langle T_{\alpha,B}(1) \rangle \sim 2^{-\alpha} \left(\frac{\alpha}{e}\right)^{\alpha/2} \sqrt{\pi\alpha}$ quand $\alpha \rightarrow \infty$. On peut retrouver sur ce graphique les valeurs $\langle T_{-1}(t=1) \rangle = 4\sqrt{\frac{2}{\pi}} \log 2 = 2.21\dots$ [85, 119] et $\langle T_{1,B}(t=1) \rangle = \sqrt{\frac{\pi}{8}} = 0.68\dots$ [148, 151].

Premier moment de la fonctionnelle du maximum du brownien libre T_α

$$\langle T_\alpha(t) \rangle = \frac{(2t)^{1+\frac{\alpha}{2}} (2 - 2^{-\alpha}) \Gamma\left(\frac{1+\alpha}{2}\right)}{(2+\alpha)\sqrt{\pi}}, \quad (5.80)$$

où Γ est la fonction Gamma (13). On a tracé en figure 5.8 ce moment vu comme une fonction de α . En étudiant la limite pour $\alpha = -1$ de cette équation, on peut obtenir la valeur de la constante c_0 (5.73) que nous avons donnée en (5.74). Dans le cas du pont, on obtient pour $\alpha \in]2, +\infty[$

Premier moment de la fonctionnelle du maximum du pont $T_{\alpha,B}$

$$\langle T_{\alpha,B}(t) \rangle = \frac{t^{1+\alpha/2}}{2^{\alpha/2}} \Gamma\left(1 + \frac{\alpha}{2}\right), \quad (5.81)$$

également tracée sur la figure 5.8. Pour $\alpha = 1$, on obtient le premier moment de la distribution d'Airy (5.79), $\langle T_{1,B}(t=1) \rangle = \sqrt{\pi/8}$. Pour $\alpha = -1$, on trouve une valeur du c_0 de l'équation (5.73) pour un pont, $c_{0,B} = T_{-1,B}(t=1)/2 = \sqrt{\pi/2} = 1.25331\dots$. Cette valeur est plus grande pour le pont que pour le brownien libre ce qui veut dire que dans ce cas, le coût moyen de l'algorithme optimal pour trouver le maximum d'une marche aléatoire est plus grand si on impose des conditions aux bords périodiques car la largeur d'un pont est typiquement plus petite que celle d'un brownien libre (on peut le

voir par exemple sur la figure 2.1 mais aussi à partir des formules (5.38) et (5.64)). Cette méthode ne permettant pas de calculer facilement les moments d'ordre supérieur, nous allons désormais appliquer le formalisme que nous avons construit à l'aide des intégrales de chemin au cas des fonctionnelles $T_\alpha(t)$ (5.69) et $T_{\alpha,B}(t)$ (5.70).

Cas du brownien libre

Dans le cas du brownien libre, nous démarrons de l'équation

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \frac{1}{2} \left(\int_0^\infty dy \partial_x \tilde{G}_s(0, y) \right)^2, \quad (5.82)$$

obtenue en combinant les équations (5.4), (5.5) et (5.15) et où $\tilde{G}_s(x, y)$ est la fonction de Green vérifiant

$$\left[-\frac{d^2}{dx^2} + \lambda V(x) + V_{\text{wall}}(x) + s \right] \tilde{G}_s(x, y) = \delta(x - y). \quad (5.83)$$

Comme nous le verrons par la suite, on peut pour certains potentiels $V(x)$ particuliers résoudre exactement cette équation. Il est néanmoins possible d'extraire de cette équation les différents moments de la fonctionnelle pour n'importe quel potentiel et donc en particulier pour $T_\alpha(t)$. En dérivant successivement par rapport à λ l'équation (5.82) évaluée en $\lambda = 0$, on obtient les différents moments de $T_\alpha(t)$. Ces dérivées successives s'expriment comme des combinaisons des dérivées de $\tilde{G}_s(x, y)$ par rapport à λ . On décompose alors $\tilde{G}_s(x, y)$ sous la forme

$$\tilde{G}_s(x, y) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} f_n(x, y), \quad (5.84)$$

on peut alors écrire

$$\int_0^\infty \langle T_\alpha^k(t) \rangle e^{-st} dt = (-1)^k \frac{k!}{2} \sum_{n=0}^k b_n b_{k-n}, \quad b_n = \frac{1}{n!} \int_0^\infty \partial_x f_n(x, y) \Big|_{x=0} dy. \quad (5.85)$$

Il suffit alors de résoudre perturbativement l'équation (5.83) pour obtenir les différentes fonctions $f_n(x, y)$ de (5.84). Au premier ordre en puissance de λ , on trouve aisément

$$f_0(x, y) = \frac{1}{\sqrt{2s}} (e^{-\sqrt{2s}|x-y|} - e^{-\sqrt{2s}(x+y)}). \quad (5.86)$$

Remarquons qu'on retrouve bien la transformée de Laplace du propagateur libre (3.9). En étudiant la relation (5.83), on trouve que les ordres suivants vérifient la relation de récurrence

$$f_n(x, y) = -n \int_0^\infty dz f_0(x, z) V(z) f_{n-1}(x, y). \quad n \geq 1. \quad (5.87)$$

que l'on peut résoudre symboliquement sous la forme

$$f_n(x, y) = (-1)^n n! \int \prod_{i=1}^n dz_i f_0(x, z_1) \prod_{i=1}^{n-1} (V(z_i) f_0(z_i, z_{i+1})) V(z_n) f_0(z_n, y). \quad (5.88)$$

En utilisant alors cette relation (5.88) combinée avec l'équation (5.85), il est possible de calculer (au moins formellement) par récursion les moments de n'importe quelle fonctionnelle du maximum pour le brownien libre. Remarquons que la formule que nous avons

obtenue est la généralisation du formalisme des propagateurs que nous avons utilisé dans la section 3 pour lequel le potentiel était $V(x) = \delta(x - r)$ ce qui permet de simplifier la formule (5.88). Calculons alors les deux premiers moments de la fonctionnelle $T_\alpha(t)$ pour laquelle le potentiel est donné par $V(x) = x^\alpha$ avec $\alpha \in]-2, \infty[$. En utilisant (5.85), le premier moment est donné par

$$\int_0^\infty dt e^{-st} \langle T_\alpha(t) \rangle = - \left(\int_0^\infty dy \partial_x f_1(x, y) \Big|_{x=0} \right) \left(\int_0^\infty dy \partial_x f_0(x, y) \Big|_{x=0} \right), \quad (5.89)$$

Où $f_0(x, y)$ est donnée par (5.86) et où la fonction $f_1(x, y)$ peut être calculée en utilisant (5.88). En remarquant que $\partial_x f_0(0, y) = 2e^{-\sqrt{2s}y}$, on peut directement calculer la quantité qui nous intéresse pour (5.89)

$$\begin{aligned} \int_0^\infty dy \partial_x f_1(0, y) &= - \int_0^\infty dy \int_0^\infty dz \partial_x f_0(0, z) z^\alpha f_0(z, y) \\ &= - \frac{2}{(2s)^{\frac{\alpha+3}{2}}} \int_0^\infty dy \int_0^\infty dz e^{-z} z^\alpha (e^{-|z-y|} - e^{-(z+y)}) \\ &= - \frac{2}{(2s)^{\frac{\alpha+3}{2}}} \Gamma(\alpha + 1) (2 - 2^{-\alpha}), \end{aligned} \quad (5.90)$$

où l'on a fait les changements de variables $z \rightarrow \sqrt{2s}z$ et $y \rightarrow \sqrt{2s}y$. Ainsi en utilisant dans l'équation (5.89) le résultat (5.90) ainsi que $\int_0^\infty dy \partial_x f_0(x, y) \Big|_{x=0} = \sqrt{2/s}$, on obtient le premier moment

$$\int_0^\infty dt e^{-st} \langle T_\alpha(t) \rangle = \frac{4}{(2s)^{\alpha/2+2}} \Gamma(\alpha + 1) (2 - 2^{-\alpha}), \quad (5.91)$$

qui après une inversion de Laplace redonne le résultat (5.80) que nous avons obtenu en utilisant la densité à partir du maximum (3.18). Pour le second moment, la formule (5.85) s'écrit

$$\begin{aligned} \int_0^\infty dt e^{-st} \langle T_\alpha^2(t) \rangle &= \left(\int_0^\infty dy \partial_x f_1(0, y) \right)^2 \\ &+ \left(\int_0^\infty dy \partial_x f_0(0, y) \right) \left(\int_0^\infty dy \partial_x f_2(0, y) \right). \end{aligned} \quad (5.92)$$

En utilisant encore une fois les relations (5.86) et (5.88) avec le changement de variables similaires $u \rightarrow \sqrt{2s}z_1, v \rightarrow \sqrt{2s}z_2$ et $y' \rightarrow \sqrt{2s}y$.

$$\begin{aligned} \int_0^\infty dy \partial_x f_2(0, y) &= 2 \int_0^\infty dy \int_0^\infty dz_1 \int_0^\infty dz_2 \partial_x f_0(0, z_1) z_1^\alpha f_0(z_1, z_2) z_2^\alpha f_0(z_2, y) \\ &= \frac{4}{(2s)^{5/2+\alpha}} \int_0^\infty dy' \int_0^\infty du \int_0^\infty dv e^{-u} u^\alpha (e^{-|u-v|} - e^{-(u+v)}) v^\alpha (e^{-|v-y'|} - e^{-(v+y')}) \\ &= \frac{8}{(2s)^{5/2+\alpha}} C_\alpha, \quad C_\alpha = \int_0^\infty du \int_0^\infty dv e^{-u} u^\alpha (e^{-|u-v|} - e^{-(u+v)}) v^\alpha (1 - e^{-v}). \end{aligned} \quad (5.93)$$

Il est possible d'exprimer la constante C_α sous la forme

$$C_\alpha = \sum_{n=1}^\infty \frac{\Gamma(2 + 2\alpha + n)}{n! 2^{2+2\alpha+n} (1 + \alpha + n)} + \frac{(4^{1+\alpha} - 1) \Gamma(3 + 2\alpha)}{2^{3+2\alpha} (1 + \alpha)^2} - \frac{2^{1+\alpha} - 1}{2^{2+2\alpha}} \Gamma(1 + \alpha)^2. \quad (5.94)$$

Finalement, en regroupant les équations (5.86), (5.90), (5.92) et (5.93), on obtient

$$\int_0^\infty dt e^{-st} \langle T_\alpha^2(t) \rangle = \frac{4}{(2s)^{3+\alpha}} \left(\Gamma(\alpha + 1)^2 (2 - 2^{-\alpha})^2 + 4C_\alpha \right), \quad (5.95)$$

qui après une transformation de Laplace inverse donne le résultat

Second moment de la fonctionnelle du maximum du brownien libre T_α

$$\begin{aligned} \langle T_\alpha^2(t) \rangle &= \frac{t^{2+\alpha}}{2^{3\alpha}\Gamma(3+\alpha)} \left(\Gamma(\alpha+1)^2(2^\alpha-1)(2^{\alpha+1}-1) \right. \\ &\quad \left. + \frac{\Gamma(3+2\alpha)(4^{\alpha+1}-1)}{4(1+\alpha)^2} + \sum_{n=1}^{\infty} \frac{\Gamma(2+2\alpha+n)}{n!2^{1+n}(1+\alpha+n)} \right) \quad (5.96) \\ &= \frac{t^{2+\alpha}}{2^{3\alpha}(\alpha+1)\Gamma(\alpha+3)} \left((2^\alpha-1)(2^{\alpha+1}-1)(\alpha+1)\Gamma(\alpha+1)^2 \right. \\ &\quad \left. + \Gamma(2\alpha+2)(4^{\alpha+1}-2^{\alpha+1}(\alpha+1)B_{\frac{1}{2}}(\alpha+2, -2(\alpha+1)) - 1) \right). \quad (5.97) \end{aligned}$$

où $B_z(a, b)$ est la fonction bêta incomplète (14). En particulier, en prenant précautionneusement la limite $\alpha \rightarrow -1$, c'est-à-dire appliquée au problème de recherche du maximum d'un marcheur aléatoire, dans l'expression (5.96), on retrouve la formule

$$\lim_{\alpha \rightarrow -1} \langle T_\alpha^2(1) \rangle = \frac{\pi^2}{3} + 4 \log(2)^2, \quad (5.98)$$

résultat que l'on peut retrouver dans [28].

Cas du pont

Dans le cas du pont, pour un potentiel $V(x)$ général, on utilise les formules (5.24) et (5.25) pour trouver

$$\langle e^{-\lambda \int_0^t d\tau V(x_{\max}-x(\tau))} \rangle = \sqrt{\frac{\pi}{2}} t^{3/2} \partial_x \partial_y G_t(x, y) \Big|_{x=y=0}, \quad (5.99)$$

que l'on peut réécrire dans le domaine de Laplace comme

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle e^{-\lambda \int_0^t d\tau V(x_{\max}-x(\tau))} \rangle = \sqrt{\frac{\pi}{2}} \partial_x \partial_y \tilde{G}_s(x, y) \Big|_{x=y=0}, \quad (5.100)$$

où $\tilde{G}_s(x, y)$ est la fonction de Green vérifiant (5.83). Via un développement en puissance de λ , on déduit directement de (5.100)

$$\int_0^\infty e^{-st} t^{-3/2} \langle \int_0^t V(x_{\max,B} - x_B(\tau)) d\tau \rangle^k dt = (-1)^k \sqrt{\frac{\pi}{2}} \partial_x \partial_y f_k(x, y) \Big|_{x=y=0}, \quad (5.101)$$

où les fonctions $f_k(x, y)$ sont toujours données par (5.88) avec la même condition initiale (5.86). Appliqué au cas particulier où $V(x) = x^\alpha$, on trouve pour le premier moment

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha,B}(t) \rangle = -\sqrt{\frac{\pi}{2}} \frac{\partial^2}{\partial x \partial y} f_1(x, y) \Big|_{x=y=0}. \quad (5.102)$$

Utilisant (5.86) et (5.88) ainsi que le changement de variable $z \rightarrow \sqrt{2s}z$, on obtient

$$\begin{aligned} \int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha,B}(t) \rangle &= \sqrt{\frac{\pi}{2}} \int_0^\infty dz \partial_x f_0(0, z) z^\alpha \partial_y f_0(z, 0) \\ &= \sqrt{\frac{\pi}{2}} \frac{4}{(2s)^{\frac{\alpha+1}{2}}} \int_0^\infty dz e^{-2z} z^\alpha = \sqrt{\frac{\pi}{2}} \frac{2^{1-\alpha}}{(2s)^{\frac{\alpha+1}{2}}} \Gamma(1+\alpha), \end{aligned} \quad (5.103)$$

qui après une inversion de Laplace redonne le résultat (5.81) que nous avons obtenu en utilisant la densité à partir du maximum (3.27).

Il est également possible de calculer le second moment $\langle T_{\alpha,B}^2(t) \rangle$ en utilisant (5.100)

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha,B}^2(t) \rangle = \sqrt{\frac{\pi}{2}} \partial_x \partial_y f_2(0,0), \quad (5.104)$$

où $f_2(x,y)$ s'obtient à partir de (5.86) et (5.88) avec les changements de variables $u \rightarrow \sqrt{2s}z_1, v \rightarrow \sqrt{2s}z_2$

$$\begin{aligned} \int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha,B}^2(t) \rangle &= \sqrt{2\pi} \int_0^\infty dz_1 \int_0^\infty dz_2 \partial_x f_0(0, z_1) z_1^\alpha f_0(z_1, z_2) z_2^\alpha \partial_y f_0(z_2, 0) \\ &= \frac{\sqrt{2\pi}}{2^{2\alpha}(2s)^{\alpha+3/2}} \left(\frac{2\Gamma(2+2\alpha)}{1+\alpha} - \Gamma(1+\alpha)^2 \right), \end{aligned} \quad (5.105)$$

qui après une inversion de Laplace devient

Second moment de la fonctionnelle du maximum du pont $T_{\alpha,B}$

$$\langle T_{\alpha,B}^2(t) \rangle = t^{\alpha+2} \frac{\sqrt{\pi}}{2^{3\alpha+1}\Gamma(\alpha + \frac{3}{2})} \left(\frac{2\Gamma(2+2\alpha)}{1+\alpha} - \Gamma(1+\alpha)^2 \right). \quad (5.106)$$

En particulier, $\langle T_{1,B}^2(t=1) \rangle = \frac{5}{12}$ qui est le second moment de la distribution d'Airy (5.79) pour le problème de l'aire sous l'excursion [105, 106]. On a tracé sur la figure 5.9 les seconds moments $\langle T_\alpha^2(t=1) \rangle$ pour le brownien libre et $\langle T_{\alpha,B}^2(t=1) \rangle$ pour le pont.

5.2.3 Cas exactement soluble du potentiel $V(x) = 1/x$

Dans cette partie, on applique nos différents résultats au cas du potentiel $V(x) = 1/x$ qui correspond donc à l'algorithme de recherche du maximum d'une marche aléatoire (5.73) [28, 119].

5.2.4 Cas du brownien libre

Dans le cas du brownien libre, la fonction de Green $G_s(x,y)$ équation (5.19) peut être écrite en fonction de $u_s(x)$ et $v_s(x)$, deux solutions indépendantes de l'équation de Schrödinger

$$\left[-\frac{1}{2} \frac{d^2}{dx^2} + \frac{\lambda}{x} + s \right] \psi(x) = 0, \quad s > 0, \quad (5.107)$$

qui possèdent les comportements asymptotiques

$$u_s(0) = 0 \quad v_s(y \rightarrow \infty) = 0. \quad (5.108)$$

Dans ce cas, cette équation peut être résolue et les solutions recherchées sont

$$u_s(x) = \mathcal{A} e^{-\sqrt{2s}x} {}_1F_1 \left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}x \right), \quad (5.109)$$

$$v_s(x) = \mathcal{B} e^{-\sqrt{2s}x} U \left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}x \right), \quad (5.110)$$

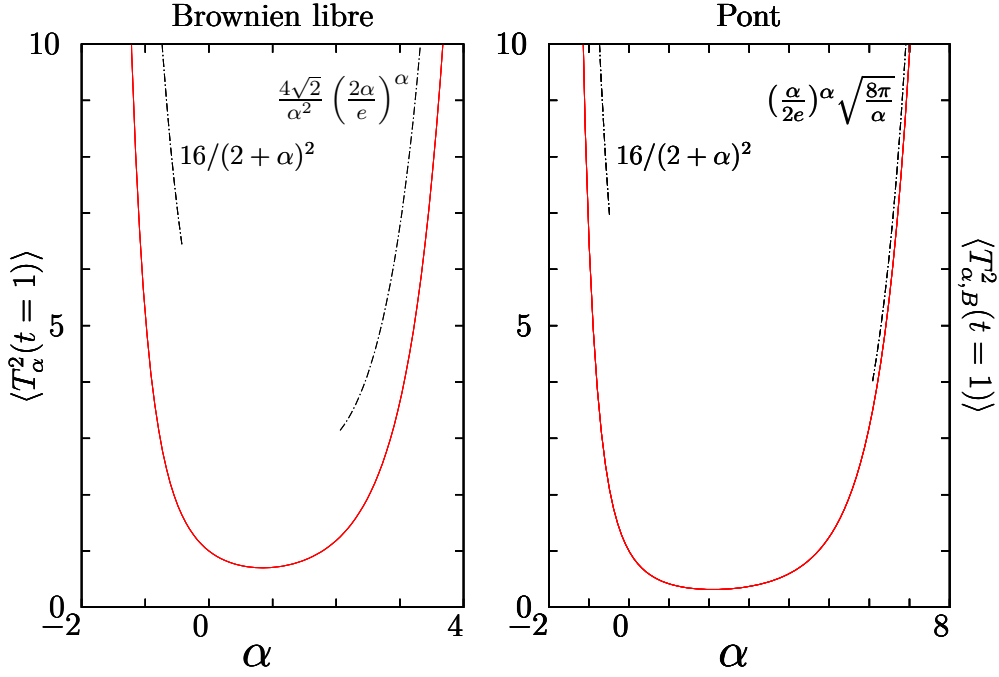


FIGURE 5.9: Seconds moments des fonctionnelles pour le brownien libre $\langle T_\alpha^2(t=1) \rangle$ équation (5.96) (à gauche) et pour le pont $\langle T_{\alpha,B}^2(t=1) \rangle$ équation (5.106) (droite) en fonction de α (à droite). Dans les deux cas, ces moments divergent comme $\sim 16/(\alpha+2)^2$ pour $\alpha \rightarrow -2$, décroissent pour atteindre une valeur minimale en $\alpha \approx 0.848$ pour le brownien libre et $\alpha \approx 2.093$ pour le pont puis croissent jusqu'à avoir les comportements asymptotiques $\langle T_\alpha^2(1) \rangle \sim \frac{4\sqrt{2}}{\alpha^2} \left(\frac{2\alpha}{e}\right)^\alpha$ et $\langle T_{\alpha,B}^2(1) \rangle \sim \left(\frac{\alpha}{2e}\right)^\alpha \sqrt{\frac{8\pi}{\alpha}}$ à grand α . On peut retrouver sur ce graphique les valeurs $\langle T_{-1}^2(t=1) \rangle = \frac{\pi^2}{3} + 4 \log(2)^2 = 4.38\dots$ [85, 119] et $\langle T_{1,B}^2(t=1) \rangle = \frac{5}{12} = 0.42\dots$ [148, 151].

où ${}_1F_1(a, b, x)$ et $U(a, b, x)$ sont respectivement les fonctions hypergéométriques de Kummer (17) et de Tricomi (18) et où \mathcal{A} et \mathcal{B} sont des constantes de normalisation dont les valeurs n'auront pas d'influence par la suite. Le wronskien W s'écrit

$$W = u'_s(x)v_s(x) - u_s(x)v'_s(x) = \frac{\mathcal{A}\mathcal{B}}{2\lambda\Gamma(\lambda/\sqrt{2s})}. \quad (5.111)$$

La fonction $\tilde{\varphi}(s)$ de l'équation (5.21) en utilisant nos résultats (5.109) et (5.110) s'écrit alors

$$\tilde{\varphi}(s) = 2^{3/2}\lambda\Gamma\left(\frac{\lambda}{\sqrt{2s}}\right) \int_0^\infty e^{-\sqrt{2s}y} y U\left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}y\right) dy, \quad (5.112)$$

que l'on peut réécrire sous la forme

$$\tilde{\varphi}(s) = \frac{1}{\sqrt{s}} G\left(\frac{\lambda}{\sqrt{2s}}\right), \quad G(x) = 2 \sum_{k=0}^\infty (-1)^{k+1} \eta(k) x^k, \quad (5.113)$$

où $\eta(k)$ est la fonction êta de Dirichlet (19). En utilisant alors le formalisme (5.4) et (5.5) combiné avec ce dernier résultat (5.113) on obtient alors

$$\int_0^\infty e^{-st} \langle e^{-\lambda \int_0^t \frac{d\tau}{x_{\max}-x(\tau)}} \rangle dt = \frac{4}{s} \sum_{n=0}^\infty (-1)^n \frac{\lambda^n}{(2s)^{n/2}} \sum_{k=0}^n \eta(k) \eta(n-k), \quad (5.114)$$

dont on peut tirer les différents moments par un développement en puissance de λ

Moments de la fonctionnelle du maximum du brownien libre T_{-1}

$$\langle T_{-1}^k(1) \rangle = \Gamma\left(\frac{k+1}{2}\right) \frac{2^{\frac{k}{2}+2}}{\sqrt{\pi}} \sum_{m=0}^k \eta(m) \eta(k-m) . \quad (5.115)$$

On retrouve les résultats obtenus avec une méthode probabiliste par Chassaing *et al.* [28] via le formalisme de l'intégrale de chemin.

Cas du pont

Dans le cas du point, nous partons de la formule (5.24) couplée avec (5.25) et appliquée à un potentiel $V(x) = 1/x$

$$\left\langle \exp \left[-\lambda \int_0^t \frac{d\tau}{x_{\max,B} - x_B(\tau)} \right] \right\rangle = \lim_{\varepsilon \rightarrow 0} \sqrt{\frac{\pi t^3}{2}} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\varepsilon^2} , \quad (5.116)$$

avec H_λ le hamiltonien (5.7) qui s'écrit dans ce cas

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dx^2} + \frac{\lambda}{x} + V_{\text{wall}}(x) , \quad \lambda \geq 0 , \quad (5.117)$$

Pour calculer (5.116), on décompose sur la base des vecteurs propres $|\phi_E\rangle$ de H_λ

$$H_\lambda |\phi_E\rangle = E |\phi_E\rangle . \quad (5.118)$$

Puisque λ est positif, le hamiltonien H_λ (5.117) ne possède aucun état localisé et son spectre est le demi-axe positif $E > 0$. À cause de la divergence du potentiel en 0, les vecteurs satisfont la condition aux limites

$$\lim_{x \rightarrow 0^+} \phi_E(x) = \lim_{x \rightarrow 0^+} \langle x | \phi_E \rangle = 0 . \quad (5.119)$$

La solution générale de (5.117) de valeurs propres E (5.118) est donnée par

$$\begin{aligned} \phi_E(x) &= c_E x e^{i\sqrt{2E}x} {}_1F_1\left(1 - i\frac{s}{\sqrt{2E}}, 2, 2i\sqrt{2E}x\right) \\ &+ d_E x e^{-i\sqrt{2E}x} U\left(1 - i\frac{s}{\sqrt{2E}}, 2, 2i\sqrt{2E}x\right) . \end{aligned} \quad (5.120)$$

La condition aux limites (5.119) impose alors $d_E = 0$. La condition de normalisation

$$\int_0^\infty \phi_{E'}^*(x) \phi_E(x) dx = \delta(E - E') . \quad (5.121)$$

permet de calculer la seconde constante

$$|c_E|^2 = \frac{4\lambda}{\exp\left(\frac{2\pi\lambda}{\sqrt{2E}}\right) - 1} . \quad (5.122)$$

En utilisant ces résultats dans l'équation (5.116), on trouve

$$\left\langle \exp \left[-\lambda \int_0^t \frac{d\tau}{x_{\max,B} - x_B(\tau)} \right] \right\rangle = 4t^{3/2} \sqrt{\frac{\pi}{2}} \lambda \int_0^\infty dE \frac{e^{-Et}}{\exp\left[2\pi\lambda/(\sqrt{2E})\right] - 1} . \quad (5.123)$$

Il est alors possible d'inverser cette transformée de Laplace à l'aide du théorème des résidus pour obtenir la densité $p_B(s)$ de $T_{-1,B}(1)$, $p_B(s)ds = \Pr . [T_{-1,B}(1) \in [s, s + ds]]$

$$p_B(s) = 2s \sum_{k=1}^{\infty} k^2 (k^2 s^2 - 3) e^{-\frac{k^2 s^2}{2}} = \frac{8\sqrt{2}\pi^{5/2}}{s^6} \sum_{k=1}^{\infty} k^2 (4\pi^2 k^2 - 3s^2) e^{-2\pi^2 \frac{k^2}{s^2}}, \quad (5.124)$$

les deux sommes étant reliées par une resommation de Poisson. On déduit de cette densité les différents moments

— Moments de la fonctionnelle du maximum du pont $T_{-1,B}$ —

$$\langle T_{-1,B}^k(1) \rangle = -2^{1+\frac{k}{2}} \pi^{k-\frac{1}{2}} k \Gamma \left(\frac{3}{2} - \frac{k}{2} \right) \zeta(1-k), \quad (5.125)$$

avec $\zeta(k)$ la fonction zêta de Riemann (19). On retrouve pour le premier et second moment les résultats que l'on peut obtenir à partir de (5.81) et (5.106) pour $\alpha = -1$.

Appendices

Annexe A

La famille de fonctions $\Phi^{(j)}$

Cette annexe introduit la famille de fonctions $\{\Phi^{(j)}, j \in \mathbb{N}\}$ satisfaisant

$$\frac{e^{-\sqrt{2s}u}}{(2s)^{\frac{j+1}{2}}} = \int_0^\infty t^{\frac{j-1}{2}} \Phi^{(j)}\left(\frac{u}{\sqrt{t}}\right) e^{-st} dt. \quad (\text{A.1})$$

De cette définition, il est facile [27] de démontrer que ces fonctions vérifient la relation de récurrence

$$\Phi^{(j+1)}(x) = \int_x^\infty \Phi^{(j)}(u) du. \quad (\text{A.2})$$

Ces fonctions peuvent donc être calculées par itération, on trouve ainsi pour les premières

$$\Phi^{(0)}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \quad (\text{A.3})$$

$$\Phi^{(1)}(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right), \quad (\text{A.4})$$

$$\Phi^{(2)}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} - \frac{1}{2} x \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right). \quad (\text{A.5})$$

avec erfc la fonction d'erreur complémentaire (11). Ces fonctions deviennent de plus en plus compliquées avec j . Plus généralement, il est possible de montrer [27] que ces fonctions peuvent toutes s'écrire sous la forme

$$\Phi^{(j)}(x) = p_j(x) \Phi^{(0)}(x) + q_j(x) \Phi^{(1)}(x), \quad (\text{A.6})$$

où $p_j(x)$ et $q_j(x)$ sont des polynômes rationnels de degré respectivement $j-2$ et $j-1$ pour $j \geq 2$ [27]. On trouvera dans [27] des algorithmes performants pour calculer ces polynômes. Remarquons le lien avec la famille de fonctions étudiée par Takács [150]

$$J_k(r) = r^{1+k} \int_1^\infty dx e^{-(xr)^2/2} (x-1)^k = \sqrt{2\pi} k! \phi^{(k+1)}(r). \quad (\text{A.7})$$

Annexe B

Divers calculs dans le cas de plusieurs ponts indépendants

Cette annexe présente plusieurs calculs dont les résultats sont discutés dans la section 4.2.1.

B.1 Calcul de la densité à partir du maximum pour n ponts indépendants

L'objectif de cette section est de calculer la densité à partir du maximum pour n ponts indépendants (4.24). Cette formule se réécrit avec la constante de normalisation (4.26)

$$\begin{aligned} \langle \rho_{\max, B, n}(r, t) \rangle &= \frac{(2\pi t)^{n/2}}{2} \int_0^\infty dM \left(2I_1(M, r, t) G_M(0|0, t)^{n-1} \right. \\ &\quad \left. + (n-1) I_2(M, t) I_3(M, r, t) G_M(0|0, t)^{n-2} \right), \end{aligned} \quad (\text{B.1})$$

où on a défini les intégrales

$$\begin{aligned} I_1(M, r, t) &= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-2} \int_0^t dt_{\max} \int_0^{t_{\max}} d\tau G_M(0|M-r, \tau) G_M(M-r|M-\varepsilon, t_{\max}-\tau) \\ &\quad \times G_M(M-\varepsilon|0, t-t_{\max}) \end{aligned} \quad (\text{B.2})$$

$$\begin{aligned} &= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-2} \int_0^t dt_{\max} \int_{t_{\max}}^t d\tau G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|M-r, \tau-t_{\max}) \\ &\quad \times G_M(M-r|0, t-\tau) \end{aligned} \quad (\text{B.3})$$

$$I_2(M, t) = \lim_{\varepsilon \rightarrow 0} \int_0^t dt_{\max} G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|0, t-t_{\max}) \quad (\text{B.4})$$

$$I_3(M, r, t) = \int_0^t d\tau G_M(0|M-r, \tau) G_M(M-r|0, t-\tau). \quad (\text{B.5})$$

Ces intégrales temporelles se calculent facilement en utilisant la transformation de Laplace et le propagateur (3.9)

$$I_1(M, r, t) = 2\sqrt{\frac{2}{\pi t}} \left(e^{-2\max(r, M)^2} - e^{-2(M+r)^2} \right), \quad (\text{B.6})$$

$$I_2(M, t) = 4\sqrt{\frac{2}{\pi t^3}} M e^{-\frac{2M^2}{t}}, \quad (\text{B.7})$$

$$I_3(M, r, t) = \Phi^{(1)}\left(2\frac{|M-r|}{\sqrt{t}}\right) + \Phi^{(1)}\left(2\frac{M+r}{\sqrt{t}}\right) - 2\Phi^{(1)}\left(2\frac{\max(M, r)}{\sqrt{t}}\right). \quad (\text{B.8})$$

On utilise ensuite le binôme de Newton pour développer le propagateur (3.8)

$$G_M(0|0, t)^k = \left(1 - e^{-2M^2}\right)^k (2\pi)^{-k/2} = \sum_{i=1}^k \binom{k}{i} (-1)^i e^{-2iM^2}. \quad (\text{B.9})$$

Il suffit ensuite de calculer l'intégrale sur la variable M dans (B.1) terme à terme pour trouver le résultat final donné en équation (4.27).

B.2 Calcul de $\langle r \rangle$ dans la limite d'un grand nombre de ponts indépendants

Pour calculer la première correction à $\langle r \rangle$, nous démarrons de la formule exacte (4.30) que nous allons étudier dans la limite asymptotique $n \rightarrow \infty$

$$\langle r \rangle = \int_0^\infty dM \left(1 - \left(1 - e^{-2M^2}\right)^n\right). \quad (\text{B.10})$$

Cherchons tout d'abord le premier terme du développement. En utilisant le changement de variable $M = u\sqrt{\frac{\ln n}{2}}$

$$\langle r \rangle = \sqrt{\frac{\ln n}{2}} \int_0^\infty du \left(1 - \left(1 - n^{-u^2}\right)^n\right). \quad (\text{B.11})$$

Or

$$1 - \left(1 - n^{-u^2}\right)^n \xrightarrow{n \rightarrow \infty} \begin{cases} 1 & \text{pour } 0 < u < 1 \\ 0 & \text{pour } u > 1, \end{cases} \quad (\text{B.12})$$

On trouve finalement que la moyenne de r au premier ordre est

$$\langle r \rangle = \sqrt{\frac{\ln n}{2}}. \quad (\text{B.13})$$

Pour analyser l'ordre suivant, nous reprenons l'intégrale (B.10) dans laquelle nous utilisons le changement de variable $M = \sqrt{\frac{\ln n}{2}} + \sqrt{\frac{2}{\ln n}}v$

$$\langle r \rangle = \sqrt{\frac{2}{\ln n}} \int_{-\frac{\ln n}{2}}^\infty dv \left(1 - \left(1 - \frac{e^{-4v(1+\frac{v}{\ln n})}}{n}\right)^n\right) \quad (\text{B.14})$$

$$= \sqrt{\frac{2}{\ln n}} \int_{-\frac{\ln n}{2}}^\infty dv \left(1 - e^{-e^{-4v}}\right) + o\left((\ln n)^{-3/2}\right) \quad (\text{B.15})$$

$$= \sqrt{\frac{\ln n}{2}} + \frac{\gamma_E}{2\sqrt{2\ln n}} + o\left((\ln n)^{-3/2}\right), \quad (\text{B.16})$$

où $\gamma_E = -\int_0^\infty e^{-t} \ln t dt = 0.577\dots$ est la constante gamma d'Euler.

B.3 Limite thermodynamique d'un grand nombre de ponts indépendants

Nous démarrons l'analyse de la densité à partir du maximum dans la limite d'un grand nombre de ponts à partir de la formule (B.1). Dans cette formule, la première partie,

caractérisant les événements où le maximum et la position à r du maximum appartiennent au même pont, est d'un ordre inférieur à la deuxième, caractérisant les événements où le maximum et la position à r du maximum appartiennent à deux ponts différent, à cause du facteur $n - 1$. Ainsi, dans la limite de grand n

$$\langle \rho_{\max, B, n}(r, t) \rangle \underset{n \rightarrow \infty}{\sim} \frac{(2\pi t)^{n/2}}{2} \int_0^\infty dM (n-1) I_2(M, t) I_3(M, r, t) G_M(0|0, t)^{n-2}. \quad (\text{B.17})$$

qui se réécrit en utilisant le résultat (B.7) et le propagateur (3.9) pour $t = 1$

$$\langle \rho_{\max, B, n}(r, 1) \rangle \underset{n \rightarrow \infty}{\sim} 4n\sqrt{2\pi} \int_0^\infty dM M e^{-2M^2} (1 - e^{-2M^2})^{n-2} I_3(M, r, 1), \quad (\text{B.18})$$

la fonction I_2 étant donnée par (B.8). De la même façon que précédemment, nous utilisons le changement de variable $M = \sqrt{\frac{\ln n}{2}} + \sqrt{\frac{2}{\ln n}}v$. Pour simplifier les écritures, nous posons $a_n = \sqrt{\frac{\ln n}{2}}$. En prenant également la limite des grands r avec $r = a_n + z$, alors

$$\lim_{n \rightarrow \infty} I_3(a_n + \frac{v}{a_n}, a_n + z, 1) = \frac{1}{2} \operatorname{erfc}(\sqrt{2}|z|). \quad (\text{B.19})$$

On trouve finalement le résultat annoncé en équation (4.36)

$$\langle \rho_{\max, B, n}(r, 1) \rangle = \sqrt{\frac{\pi}{2}} \operatorname{erfc}(\sqrt{2}|z|). \quad (\text{B.20})$$

Annexe C

Cas de plusieurs ponts répulsifs

Cette annexe discute des ponts répulsifs introduits en 4.2.2. Nous présentons dans un premier temps le mouvement brownien de Dyson qui est la méthode utilisée pour générer les ponts répulsifs. Nous exposons ensuite le calcul du temps local dans la limite d'un grand nombre de ponts répulsifs.

C.1 Génération de ponts répulsifs : mouvement brownien de Dyson

Pour générer des trajectoires de ponts répulsifs, on utilise le mouvement brownien de Dyson [46]. Nous générons ainsi n^2 marches aléatoires indépendantes de N pas chacune et contraintes à revenir à l'origine en utilisant (2.2) et (2.3). Nous notons ces n^2 marches $X_j^{k,l}$ où k et l sont des indices parcourant $[1, n]$, $j \in [0, N]$ étant le nombre de sauts effectués. On construit alors les N matrices hermitiennes H_j de taille $n \times n$

$$H_j^{k,l} = \begin{cases} X_j^{k,k} & \text{si } k = l, \\ \frac{1}{\sqrt{2}} \left(X_j^{k,l} + iX_j^{l,k} \right) & \text{si } k < l, \\ \frac{1}{\sqrt{2}} \left(X_j^{l,k} - iX_j^{k,l} \right) & \text{si } k > l. \end{cases} \quad (\text{C.1})$$

Les n valeurs propres de la matrice H_j $\lambda_{i,j}$ pour $i \in [1, n]$ après renormalisation dans la limite d'un grand nombre de pas ont la même loi que des ponts répulsifs sur $\tau \in [0, 1]$ [46, 140]

$$\frac{\lambda_{i, \lfloor \tau N \rfloor}}{\sqrt{N}} \xrightarrow{N \rightarrow \infty} x_i(\tau). \quad (\text{C.2})$$

Nous avons utilisé cet algorithme pour générer les ponts répulsifs dans les différentes figures en utilisant un nombre de sauts $N = 10^4$.

C.2 Temps local des ponts répulsifs dans la limite thermodynamique de grand n

Cette section utilise la correspondance entre les marcheurs répulsifs et la théorie des matrices aléatoires gaussiennes unitaires pour calculer le temps local de n ponts répulsifs. On utilise pour cela des résultats introduits dans la partie suivante sur les matrices

aléatoires. Nous cherchons ainsi à calculer le temps local

$$\langle T_{loc,B,n}^{Vicious}(r, 1) \rangle = \frac{1}{n} \sum_{i=1}^n \langle \int_0^1 d\tau \delta(r - x_{i,B}(\tau)) \rangle, \quad (C.3)$$

où $x_{i,B}$ sont les positions des ponts répulsifs et sont donc distribués, à τ fixé, comme $\lambda_i / \sqrt{2\tau(1-\tau)}$ avec λ_i les valeurs propres d'une matrice gaussienne hermitienne $n \times n$. En utilisant la densité moyenne de valeurs propres de ce modèle $\rho_0(\lambda, n)$ (6.7), on peut donc réécrire

$$\langle T_{loc,B,n}^{Vicious}(r, 1) \rangle = \int_0^1 d\tau \frac{1}{\sqrt{2\tau(1-\tau)}} \rho_0 \left(\frac{r}{\sqrt{2\tau(1-\tau)}}, n \right). \quad (C.4)$$

En utilisant le comportement asymptotique de cette densité (6.16) et (6.12) qui converge vers la loi du demi-cercle de Wigner obligeant $r \leq \sqrt{n}$ pour avoir un résultat non nul

$$\langle T_{loc,B,n}^{Vicious}(r, 1) \rangle \underset{n \rightarrow \infty}{\sim} \int_{\frac{1}{2} \left(1 - \sqrt{1 - \frac{r^2}{n}} \right)}^{\frac{1}{2} \left(1 + \sqrt{1 - \frac{r^2}{n}} \right)} d\tau \frac{1}{\pi \sqrt{n\tau(1-\tau)}} \sqrt{1 - \frac{r^2}{4n\tau(1-\tau)}}. \quad (C.5)$$

En utilisant le changement de variable $y = 2\tau - 1$, on peut réécrire ce comportement comme

$$\lim_{n \rightarrow \infty} \sqrt{n} \langle T_{loc,B,n}^{Vicious}(\sqrt{n}x, 1) \rangle = \frac{2}{\pi} \int_0^{\sqrt{1-x^2}} dy \frac{1}{1-y^2} \sqrt{1 - \frac{x^2}{1-y^2}}. \quad (C.6)$$

En posant alors $\cos(\alpha) = x$, $\alpha \in [0, \pi/2]$ et avec le changement de variable $y = \sin(\theta)$

$$\lim_{n \rightarrow \infty} \sqrt{n} \langle T_{loc,B,n}^{Vicious}(\sqrt{n} \cos(\alpha), 1) \rangle = \frac{2}{\pi} \int_0^\alpha d\theta \sqrt{1 - \frac{\cos^2(\alpha)}{\cos^2(\theta)}}, \quad (C.7)$$

dont une primitive est

$$G(\theta) = \frac{2}{\pi} \left(\text{Arcsin} \left(\frac{\sin \theta}{\sin \alpha} \right) - \cos \alpha \text{Arctan} \left(\frac{\sqrt{2} \sin \theta \cos \alpha}{\sqrt{\cos^2 \theta - \cos^2 \alpha}} \right) \right). \quad (C.8)$$

Ainsi finalement, on trouve le comportement du temps local dans la limite de thermodynamique (et dans le régime de volume)

$$\lim_{n \rightarrow \infty} \sqrt{n} \langle T_{loc,B,n}^{Vicious}(\sqrt{n}x, 1) \rangle = \begin{cases} \lim_{\theta \rightarrow \alpha} G(\theta) - G(0) = 1 - \cos(\alpha) = 1 - |x| & \text{pour } |x| < 1 \\ 0 & \text{sinon,} \end{cases} \quad (C.9)$$

ce qui conduit au résultat (4.45).

Annexe D

Moments de la densité à partir du maximum du pont

Dans cette annexe, nous démontrons que le résultat (5.61) obtenue avec la méthode de l'intégrale de chemin permet de retrouver les moments de la densité à partir du maximum (3.25) que nous avons obtenu avec la méthode des propagateurs. Nous démarrons pour cela de la formule (5.61) qui donne la transformée de Laplace de la loi de la densité à partir du maximum

$$\langle e^{-\lambda \rho_{\max, B}(r, t)} \rangle = \sqrt{\frac{2t^3}{\pi}} \int_0^\infty \frac{k^2 e^{-\frac{k^2 t}{2}}}{1 + \frac{4\lambda}{k} \sin(kr) \left(\frac{\lambda}{k} \sin(kr) + \cos(kr) \right)} dk. \quad (D.1)$$

Nous allons obtenir les différents moments de la densité à partir du maximum $\langle \rho_{\max, B}^m(r, t) \rangle$ en développant cette expression en puissance de λ . On utilise pour cela l'égalité

$$\frac{1}{1 + \frac{4\lambda}{k} \sin(kr) \left(\frac{\lambda}{k} \sin(kr) + \cos(kr) \right)} = \frac{1}{\sin^2(kr) \left(2\frac{\lambda}{k} + \cotan(kr) \right)^2} \quad (D.2)$$

$$= \sum_{m=0}^{\infty} \frac{(-2\lambda)^m}{k^m} \sin^{m-1}(kr) \sin((m+1)kr). \quad (D.3)$$

Ainsi,

$$\langle \rho_{\max, B}^m(r, t) \rangle = m! 2^m \sqrt{\frac{2t^3}{\pi}} \int_0^\infty k^{2-m} e^{-\frac{k^2 t}{2}} \sin^{m-1}(kr) \sin((m+1)kr) \quad (D.4)$$

$$= t^{m/2} \frac{m! 2^m}{\sqrt{2\pi}} \int_{-\infty}^\infty k^{2-m} e^{-\frac{k^2 t}{2}} \sin^{m-1}(kr) \sin((m+1)kr). \quad (D.5)$$

On injecte alors l'égalité

$$k^{m-2} = \frac{-1}{(m-3)!} \int_0^\infty du u^{m-3} \cos\left(ku - m\frac{\pi}{2}\right). \quad (D.6)$$

Si m est pair, on peut développer la puissance \sin^{m-1} en utilisant

$$\sin^{m-1}(kr) = 2^{1-m} \sum_{p=0}^{m-1} (-1)^{m/2-1-p} \binom{m-1}{p} \sin((m-1-2p)kr). \quad (D.7)$$

On a de plus dans ce cas $\cos\left(ku - m\frac{\pi}{2}\right) = (-1)^{m/2} \cos(ku)$. En linéarisant le produit de fonction trigonométrique

$$\begin{aligned} & 4 \cos(ku) \sin((m-1-2p)kr) \sin((m+1)kr) \\ &= \sum_{\varsigma=\pm 1} (\cos(k(2(p+1)r + \varsigma u)) - \cos(k(2(m-p)r + \varsigma u))) , \end{aligned} \quad (D.8)$$

et en utilisant l'intégrale

$$\int_{-\infty}^{\infty} dk e^{-\frac{k^2}{2}} \cos(kx) = \sqrt{2\pi} e^{x^2/2}, \quad (\text{D.9})$$

On trouve alors

$$\begin{aligned} \langle \rho_{\max, B}^m(r, 1) \rangle &= \frac{m!}{2(m-3)!} \sum_{p=0}^{m-1} \binom{m-1}{p} (-1)^p \\ &\quad \int_0^\infty du u^{m-3} \left(e^{\frac{(2(p+1)r-u)^2}{2}} - e^{\frac{(2(m-p)r-u)^2}{2}} + e^{\frac{(2(p+1)r+u)^2}{2}} - e^{\frac{(2(m-p)r+u)^2}{2}} \right). \end{aligned} \quad (\text{D.10})$$

En utilisant la formule

$$\int_0^\infty du u^{m-3} \left(e^{\frac{(a-u)^2}{2}} + e^{\frac{(a+u)^2}{2}} \right) = (m-3)! \sqrt{2\pi} \left(\Phi^{(m-2)}(a) + \Phi^{(m-2)}(-a) \right), \quad (\text{D.11})$$

on trouve après resommation

$$\langle \rho_{\max, B}^m(r, 1) \rangle = m! \sqrt{2\pi} \sum_{p=0}^{m-1} \binom{m-1}{p} (-1)^p \left(\Phi^{(m-2)}(2(p+1)r) + \Phi^{(m-2)}(-2(p+1)r) \right), \quad (\text{D.12})$$

Par récurrence, il est facile de voir que $\Phi^{(n)}(x) + (-1)^{n+1} \Phi^{(n)}(x)$ est un polynôme de degré $n-1$ de parité $n-1$ ce qui permet de montrer que les deux sommes de l'équation (D.12) sont identiques et on retrouve ainsi les résultats obtenus avec la méthode des propagateurs (3.25).

Annexe E

Recherche en environnement inconnu : Algorithme d'Odlyzko

Dans cet appendice, nous expliquons pourquoi la fonctionnelle I (5.73) apparaît comme le coût optimal de recherche du maximum d'une marche aléatoire puis donnons, dans les grandes lignes, l'algorithme optimal d'Odlyzko [119]. Considérons la marche aléatoire

$$\begin{cases} X_0 = 0 \\ X_i = X_{i-1} + \eta_i, i \in [1, n], \end{cases} \quad (\text{E.1})$$

où les η_i sont des variables aléatoires $\{-1, +1\}$ chacun avec probabilité $1/2$. Nous recherchons alors la méthode optimale pour trouver le maximum de cette marche aléatoire $M_n = \max\{X_i, 0 \leq i \leq n\}$ en minimisant le nombre de positions X_k révélées pour trouver ce maximum. Nous nous autorisons à faire autant de calculs que nécessaires avec les positions X_k examinées, seul compte le nombre de positions révélées dont nous avons eu besoin pour trouver le maximum. Pour être performant, les algorithmes utilisent la forte corrélation qui existe entre les positions X_i . Ainsi, par exemple, si on a révélé les deux positions X_m et X_{m+k} , il est possible d'en déduire de l'information sur la trajectoire entre ces deux positions. En effet, en notant k_+ le nombre de sauts $+1$ et k_- le nombre de sauts -1 entre les positions m et $m+k$, alors

$$k_+ + k_- = k \quad (\text{E.2})$$

$$X_m + k_+ - k_- = X_{m+k}, \quad (\text{E.3})$$

et donc

$$k_+ = \frac{X_{m+k} - X_m + k}{2} \quad (\text{E.4})$$

$$k_- = \frac{X_m - X_{m+k} + k}{2}. \quad (\text{E.5})$$

Ce qui oblige la trajectoire à être plus petite que $X_m + k_+ = (X_m + X_{m+k} + k)/2$ entre m et $m+k$, comme le montre la figure 5.4 de la section 5.2.2, ce maximum étant atteint que si la trajectoire entre les positions m et $m+k$ est constituée des k_+ sauts $+1$ puis des k_- sauts -1 . Cette idée est à la base de l'algorithme u que nous avons illustré sur la figure 5.5 de la section 5.2.2 qui analyse à chaque étape le plus grand maximum des trajectoires possibles passant par les positions sondées et révèle la position de ce maximum. Le coût $C(a)$ d'un algorithme $a \in A_n$ est une variable aléatoire qui varie d'une trajectoire à l'autre, on note sa moyenne sur les différentes trajectoires possibles $\langle C(a) \rangle$. Pour comprendre pourquoi le coût optimal fait intervenir la fonctionnelle I (5.73), nous utilisons le raisonnement décrit

par Chassaing *et al.* [28]. Considérons un marcheur se déplaçant sur une ligne dont la position est notée y . Supposons que sa vitesse $v(y)$ ne peut excéder une certaine vitesse limite $z(y)$: $0 < v(y) \leq z(y)$. Le temps pour parcourir le trajet de l'origine à un point x est alors borné

$$t = \int_0^x \frac{dy}{v(y)} \geq \int_0^x \frac{dy}{z(y)} . \quad (\text{E.6})$$

Revenons maintenant à notre problème en considérant un algorithme a de coût $C(a)$ pour une certaine trajectoire. On dénote par $m_1, \dots, m_{C(a)}$ les positions que l'algorithme a révélées pour trouver le maximum M_n . Pour être sûr qu'il n'existe aucun point plus haut que la maximum dans l'intervalle $[m_i, m_{i+1}]$, il faut donc que $(X_{m_i} + X_{m_{i+1}} + m_{i+1} - m_i)/2 \leq M_n$ (comme le montre la figure 5.4 de la section 5.2.2) et donc

$$m_{i+1} - m_i \leq 2M_n - X_{m_i} - X_{m_{i+1}} . \quad (\text{E.7})$$

Dans la limite de grand n , en utilisant le fait que les marches aléatoires sont, comme nous allons le voir en équation (E.9), presque toutes à "variations lentes" [119], on a $2M_n - X_{m_i} - X_{m_{i+1}} \sim 2(M_n - X_{m_i})$. En faisant correspondre $m_{i+1} - m_i$ à la vitesse de l'algorithme $v(m_i)$ au point m_i , l'inégalité (E.7) fait apparaître une vitesse limite et donc en utilisant l'analogie (E.6), le coût de l'algorithme $C(a)$ est borné par

$$C(a) \geq \frac{1}{2} \sum_{k=1}^n \frac{1}{M_n - X_k} , \quad (\text{E.8})$$

qui dans la limite continue de grand n fait intervenir la fonctionnelle du maximum I (5.73) qui apparaît donc comme le coût optimal.

Algorithme optimal d'Odlyzko

Nous décrivons ici plus en détail l'algorithme optimal proposé par Odlyzko [119]. On se donne c un nombre positif très grand. La base de l'algorithme est que la plupart des marches aléatoires (E.1) sont à "variations lentes" c'est-à-dire vérifient

$$|X_{i+k} - X_i| \leq c\sqrt{k \log n} , \quad \forall i, k \text{ avec } i + k \leq n . \quad (\text{E.9})$$

Plus précisément, pour c suffisamment grand, la probabilité qu'une marche aléatoire (E.1) viole l'inégalité (E.9) décroît comme n^{-1} [28, 119] comme on peut le montrer en utilisant que $\Pr(|X_j| > x) \leq 2 \exp(-x^2/(2j))$. Ainsi, même si trouver le maximum d'une telle marche nécessite de révéler beaucoup de positions, la contribution au coût moyen restera négligeable car cela arrive avec une probabilité très faible. Au contraire, comme nous allons le voir, il est relativement simple de trouver le maximum des marches aléatoires à variations lentes (E.9).

L'algorithme d'Odlyzko est constitué de deux étapes. Notons que l'algorithme test à chaque révélation de position l'inégalité (E.9), s'il s'avère que celle-ci est violée, on arrête cette méthode et révèle toutes les positions pour trouver le maximum. Ce cas arrivant suffisamment rarement, sa contribution est négligeable.

- En premier lieu, on cherche une bonne estimation M^* du maximum M_n (que l'on ne connaît pas). Pour cela, on scanne toutes les positions $X_N, X_{2N}, X_{3N}, \dots$ avec $N = \lfloor \sqrt{n \log n} \rfloor + 1$. On note alors $M' = \max\{X_0, X_N, X_{2N}, X_{3N}, \dots\} \leq M_n$ le maximum de ces positions révélées. Si la marche est bien à variation lente (E.9), on a

$$M_n - M' \leq c\sqrt{N \log n} \leq cn^{1/4} \log n . \quad (\text{E.10})$$

Cette estimation M' du maximum M_n n'est pas encore suffisante, il est nécessaire de rescanner le voisinage des plus grandes positions X_{rN} avec un pas plus faible. Si pour un certain r on a

$$X_{rN} \geq M' - cn^{1/4} \log n, \quad (\text{E.11})$$

alors on sonde son voisinage $X_{rN \pm jK}$, $j = 1, 2, \dots \lfloor N/K \rfloor$ avec le pas $K = \lfloor n^{1/4} \rfloor + 1$. Si la marche aléatoire est à variation lente (E.9), alors le maximum M^* de toutes les positions sondées vérifie

$$0 \leq M_n - M^* \leq c\sqrt{n^{1/4} \log n} \leq n^{1/6}. \quad (\text{E.12})$$

On peut montrer [119] que la recherche de cette estimation a nécessité de révéler en moyenne $\mathcal{O}(\sqrt{n}/\log n)$ positions ce qui est négligeable devant le coût final de l'algorithme.

• À l'aide de cette estimation M^* du vrai maximum M_n , nous allons en révélant $\mathcal{O}(\sqrt{n})$ positions trouver le maximum M_n . Pour ça, nous scannons la marche entière de gauche à droite. On introduit m l'indice de la dernière position X_m que l'algorithme révèle. L'algorithme démarre en $m = 0$. On note également $M^\#$ la plus grande position révélée au temps considéré, incluant l'estimation M^* . À chaque phase de l'algorithme deux cas peuvent se produire :

- (i) si $M^\# - X_m \leq n^{1/6}$, l'algorithme dans la prochaine phase observera ce qui se passe juste à droite de la position X_m : $m \rightarrow m + 1$.
- (ii) si $M^\# - X_m > n^{1/6}$, on se situe donc dans une portion de la trajectoire très loin du maximum puisque $M_n - M^\# \leq n^{1/6}$. Il n'est donc pas nécessaire de considérer ce qu'il se passe dans le voisinage de m . En effet, en utilisant encore une fois le raisonnement de la figure 5.4, le maximum ne peut être dans l'intervalle $[m, m+k]$ avec

$$k \leq 2(M^\# - X_m) + (X_m - X_{m+k}). \quad (\text{E.13})$$

le premier terme $2(M^\# - X_m)$ étant supérieur à $2n^{1/6}$, tandis que le second $(X_m - X_{m+k})$ est, pour une marche à variation lente (E.9), plus petit que $c\sqrt{k \log n}$, on choisit de prendre k légèrement plus petit que $2(M^\# - X_m)$. Si $m+k > n$, on examine la position X_n et arrête l'algorithme. Quand le chemin complet a été scanné, le maximum M_n a été trouvé.

Pour une marche aléatoire à variation lente (E.9), on peut montrer [119] que la contribution majeure au coût de l'algorithme $\mathcal{O}(\sqrt{n})$ provient des cas (ii) quand $M^\# - X_m > n^{1/6}$. En effet, à partir d'une estimation $M_n - M^* < n^\alpha$, il est possible de montrer que la contribution des cas (i) au coût de l'algorithme est d'ordre $\mathcal{O}(n^{2\alpha})$ et est donc bien d'un ordre plus faible dès que $2\alpha < 1/2$ d'où le choix $\alpha = 1/6$ (E.12) fait par Odlyzko [119]. Puisqu'il a été nécessaire de révéler uniquement une position entre m et $m+k$ pour contrôler que les k positions sont inférieurs à l'estimation du maximum $M^\#$, k peut être interprété comme la vitesse de l'algorithme qui est légèrement plus petite que $2(M_n - X_m)$. En utilisant (E.6), le coût de l'algorithme est, au premier ordre quand n tend vers l'infini,

$$\langle C(\text{Od}) \rangle = \frac{1}{2} \left\langle \sum_{i=0}^n \frac{1}{M_n - X_i + 1} \right\rangle. \quad (\text{E.14})$$

Quand n tend vers l'infini, la marche aléatoire (E.1) converge vers le brownien libre et on retrouve donc la fonctionnelle I (5.73)

$$\frac{C(\text{Od})}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} I = \frac{1}{2} \int_0^1 \frac{d\tau}{x_{\max} - x(\tau)}. \quad (\text{E.15})$$

Remarquons que tous les arguments utilisés se transposent facilement au cas des marches aléatoires contraintes à revenir à l'origine après n pas $X_{0,B} = X_{n,B} = 0$. Dans ce cas, la marche aléatoire converge vers le pont et le coût de l'algorithme optimal, donné par la méthode d'Odlyzko, est la fonctionnelle $(1/2)T_{\alpha=-1}^B(t)$ définie en (5.70). Notons enfin que dans son papier, Odlyzko [119] utilise la probabilité que la marche de n pas (E.1) soit à une distance q du maximum M_n en X_k

$$P(n, k, q) = \Pr . [M_n - X_k = q] , \quad (\text{E.16})$$

qu'il calcule

$$P(n, k, q) = \frac{1}{2^n} \left(\binom{k}{\lfloor \frac{k+q+1}{2} \rfloor} \sum_{j=0}^q \binom{n-k}{\lfloor \frac{n-k+j}{2} \rfloor} + \binom{n-k}{\lfloor \frac{n-k-q}{2} \rfloor} \sum_{j=0}^{q-1} \binom{k}{j + \lfloor \frac{k-q+1}{2} \rfloor} \right). \quad (\text{E.17})$$

Odlyzko [119] utilise en particulier dans sa démonstration la moyenne qu'il note $E(D_q)$

$$E(D_q) = \sum_{k=0}^n P(n, k, q) , \quad (\text{E.18})$$

qui est le nombre moyen de sites visités par la marche à une distance q du maximum. Dans la limite de grand n , cette quantité est reliée à la densité à partir du maximum

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E(D_{q=\sqrt{nr}}) = \langle \rho_{\max}(r, 1) \rangle . \quad (\text{E.19})$$

Deuxième partie

Matrices Aléatoires

Chapitre 6

Introduction

6.1 Les matrices aléatoires

Le modèle des matrices aléatoires a été introduit par Wigner [163] pour modéliser les hamiltoniens très complexes des gros noyaux atomiques. Ces matrices sont constituées d'éléments qui sont eux-mêmes des variables aléatoires (pouvant être corrélées entre elles). Le noyau est ainsi vu comme quelque chose de très compliqué constitué d'un grand nombre de composants en interaction, de sorte que les détails ne devraient pas avoir d'importance. Cette modélisation est ainsi l'analogie de la physique statistique utilisée pour modéliser un gaz de particules dont on ne peut décrire précisément les vitesses et positions de chacune des particules. Désormais, le modèle des matrices aléatoires possède diverses applications dans le domaine de la physique théorique en particulier à la physique nucléaire, au chaos quantique, aux systèmes désordonnés et à la théorie des nombres [2, 113]. Nous définissons dans ce chapitre l'ensemble de matrices aléatoires gaussien unitaire (et plus généralement les ensembles gaussiens). Nous discuterons ensuite de quelques résultats sur les matrices aléatoires en particulier sur la densité de valeurs propres dans la section 6.3 et sur les valeurs propres extrémales dans la section 6.4. Mentionnons les ouvrages de référence sur la théorie des matrices aléatoires [2, 65, 113] bien plus complet que cette introduction.

6.2 Les ensembles gaussiens

L'ensemble gaussien unitaire (**GUE** pour *Gaussian Unitary Ensemble*) correspond aux matrices M carrées hermitiennes de taille $N \times N$ dont les coefficients sont distribués suivant des gaussiennes complexes indépendantes

$$M_{n,m} = \begin{cases} \mathcal{N}[0, \frac{1}{\sqrt{2}}] & \text{si } n = m, \\ \mathcal{N}[0, \frac{1}{2}] + i\mathcal{N}[0, \frac{1}{2}] & \text{si } n < m, \\ M_{m,n}^* & \text{si } n > m, \end{cases} \quad (6.1)$$

où on a noté $\mathcal{N}[m, \sigma]$ la loi normale de moyenne m et d'écart-type σ . Ce choix permet à la fois d'avoir des coefficients indépendants et d'avoir invariance par changement de base unitaire : la probabilité de M est la même que celle de UMU^\dagger avec U une matrice unitaire. Cette propriété d'invariance montre, en diagonalisant la matrice M , que l'on peut finalement oublier la base de vecteurs propres, qui sont distribués uniformément, et ne garder que la statistique des valeurs propres. Dans le cas des matrices gaussiennes unitaires, la loi jointe de densité de probabilité des N valeurs propres $(\lambda_1, \lambda_2, \dots, \lambda_N)$ est

donnée par [163]

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j} (\lambda_i - \lambda_j)^2 \exp \left(- \sum_{i=1}^N \lambda_i^2 \right), \quad (6.2)$$

avec la constante de normalisation $Z_N = 2^{-\frac{N^2}{2}} (2\pi)^{\frac{N}{2}} \prod_{j=1}^N j!$. Les N valeurs propres sont des variables aléatoires fortement corrélées qui se repoussent entre elles à cause du déterminant de Vandermonde $\prod_{i < j} (\lambda_i - \lambda_j)$. On dénotera par la suite $\langle X \rangle$ la moyenne sur les configurations des valeurs propres

$$\langle X \rangle = \int_{-\infty}^{\infty} \prod_{i=1}^N d\lambda_i X(\lambda_1, \lambda_2, \dots, \lambda_N) P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N). \quad (6.3)$$

Au lieu de considérer l'ensemble gaussien unitaire des matrices complexes hermitiennes (6.1), on peut également étudier l'ensemble gaussien orthogonal construit avec des matrices réelles symétriques (**GOE** pour *Gaussian Orthogonal Ensemble*) et l'ensemble gaussien symplectique construit avec des matrices quaternioniques self-duales (**GSE** pour *Gaussian Symplectic Ensemble*). Ces trois ensembles de matrices font partie de la famille des ensembles gaussiens β . Dans ce cas, la loi jointe (6.2) est remplacée par [163]

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_{N,\beta}} \prod_{i < j} (\lambda_i - \lambda_j)^\beta \exp \left(- \frac{\beta}{2} \sum_{i=1}^N \lambda_i^2 \right), \quad (6.4)$$

où la constante de normalisation peut être évaluée par une intégrale de Selberg [113]

$$Z_{N,\beta} = \frac{(2\pi)^{\frac{N}{2}}}{\beta^{\frac{N}{2} + \beta \frac{N(N-1)}{4}}} \frac{\prod_{j=1}^N \Gamma(1 + \beta j/2)}{\Gamma(1 + \beta/2)^N}. \quad (6.5)$$

Le paramètre β est appelé l'indice de Dyson. Celui-ci dépend de l'ensemble considéré : $\beta = 1$ pour l'ensemble gaussien orthogonal, $\beta = 2$ pour l'ensemble gaussien unitaire et $\beta = 4$ pour l'ensemble gaussien symplectique. On peut néanmoins considérer la distribution (6.4) pour un $\beta > 0$ quelconque. Dumitriu et Edelman [44] ont trouvé un modèle, que nous discuterons dans la section 7.4.1, de matrices symétriques réelles tridiagonales qui a pour distribution jointe de valeurs propres la loi (6.4) pour n'importe quel $\beta > 0$. La loi jointe (6.4) peut-être réécrite comme un poids de Boltzmann

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) \propto \exp \left(- \frac{\beta}{2} \left(\sum_{i=1}^N \lambda_i^2 - \sum_{i \neq j} \ln(\lambda_i - \lambda_j) \right) \right), \quad (6.6)$$

l'indice de Dyson β peut donc être assimilé à une température inverse. Cette interprétation amène naturellement à considérer la collection des N valeurs propres $(\lambda_1, \lambda_2, \dots, \lambda_N)$ comme les positions d'un gaz de Coulomb bidimensionnel (potentiel logarithmique) confiné par un potentiel externe quadratique et contraint à rester sur une ligne.

6.3 Densité des valeurs propres

On introduit la densité moyenne de valeurs propres

$$\rho_0(\lambda, N) = \frac{1}{N} \sum_{i=1}^N \langle \delta(\lambda_i - \lambda) \rangle. \quad (6.7)$$

La probabilité pour une valeur propre de se trouver dans l'intervalle $[\lambda, \lambda + d\lambda]$ est donnée par $\rho_0(\lambda, N)d\lambda$. Cette densité, normalisée à 1, est la distribution marginale d'une valeur propre

$$\rho_0(\lambda, N) = \int_{-\infty}^{\infty} \prod_{i=2}^N d\lambda_i P_{\text{joint}}(\lambda, \lambda_2, \dots, \lambda_N). \quad (6.8)$$

Dans la limite N grand, en supposant que le comportement typique des valeurs propres suit une loi de puissance : $\lambda_{\text{typ}} \sim N^a$, une comparaison en ordre de grandeur des deux termes de (6.6) permet de trouver le comportement typique des valeurs propres [117]

$$\sum_{i=1}^N \lambda_i^2 \sim N \lambda_{\text{typ}}^2 \sim N^{2a+1} \quad (6.9)$$

$$\sum_{i \neq j} \ln(\lambda_i - \lambda_j) \sim N^2, \quad (6.10)$$

les deux termes s'équilibrent donc quand $a = 1/2$. le comportement typique des valeurs propres est donc $\lambda_{\text{typ}} \sim \sqrt{N}$. Remarquons que cet effet provient de la répulsion entre les valeurs propres à cause du déterminant de Vandermonde $\prod_{i < j} (\lambda_i - \lambda_j)$ présent dans la loi jointe (6.4). Si les valeurs propres étaient indépendantes, on aurait λ_{typ} d'ordre 1. La densité (6.8) doit donc avoir le comportement

$$\rho_0(x, N) \underset{N \rightarrow \infty}{\sim} \frac{1}{\sqrt{N}} \rho_{0,B} \left(\frac{x}{\sqrt{N}} \right), \quad (6.11)$$

où ce régime est dit de volume (*Bulk*) et est visible lorsque l'on considère des valeurs propres typiques d'ordre \sqrt{N} . Dans ce régime, $\rho_{0,B}(x)$ est la loi du demi-cercle de Wigner [65, 113] indépendamment de β

$$\rho_{0,B}(x) = \rho_W(x) = \frac{1}{\pi} \sqrt{2 - x^2}, \quad (6.12)$$

le support des valeurs propres est alors l'intervalle fermé $[-\sqrt{2N}, \sqrt{2N}]$. Ainsi, la valeur propre maximale est proche de $\sqrt{2N}$. On peut montrer qu'il existe également un autre régime [64, 65, 113] dit de bord (*Edge*) et est justement observable lorsque l'on s'intéresse à des valeurs propres proches de la valeur propre maximale. Pour connaître l'amplitude typique des fluctuations de la valeur propre maximale autour de la valeur $\sqrt{2N}$, on utilise l'analogie de la formule (1.3)

$$\frac{1}{N} = \int_{\lambda_{\max}}^{\infty} d\lambda \rho_0(\lambda, N) \sim \frac{1}{\sqrt{N}} \int_{\lambda_{\max}}^{\sqrt{2N}} d\lambda \rho_{0,B} \left(\frac{\lambda}{\sqrt{N}} \right) \Rightarrow \lambda_{\max} - \sqrt{2N} \sim N^{-1/6}, \quad (6.13)$$

où l'on a utilisé le comportement en racine de $\rho_{0,B}$ proche du bord. Ainsi, dans le régime $|\lambda - \sqrt{2N}| = \mathcal{O}(N^{-1/6})$, on s'attend à observer la fonction d'échelle de bord [16]

$$\lim_{N \rightarrow \infty} \frac{N^{5/6}}{\sqrt{2}} \rho_0 \left(\sqrt{2N} + \frac{x}{\sqrt{2N^{1/6}}}, N \right) = \rho_{0,E}(x), \quad (6.14)$$

où $\rho_{0,E}(x)$ dépend de β et n'est connue que pour les cas $\beta = 1, 2$ et 4. Elle est donnée dans le cas $\beta = 2$ par le noyau d'Airy au même point et par des combinaisons de fonction d'Airy pour les cas $\beta = 1$ et 4 [16, 64]

$$\rho_{0,E}(x) \sim \begin{cases} [\text{Ai}'(x)]^2 - x \text{Ai}^2(x) + \frac{1}{2} \text{Ai}(x) (1 - \int_x^{\infty} dt \text{Ai}(t)), & \beta = 1, \\ [\text{Ai}'(x)]^2 - x \text{Ai}^2(x), & \beta = 2, \\ \kappa^{-1/2} \left([\text{Ai}'(\kappa x)]^2 - \kappa x \text{Ai}^2(\kappa x) - \frac{1}{2} \text{Ai}(\kappa x) \int_{\kappa x}^{\infty} dt \text{Ai}(t) \right), & \beta = 4, \end{cases} \quad (6.15)$$

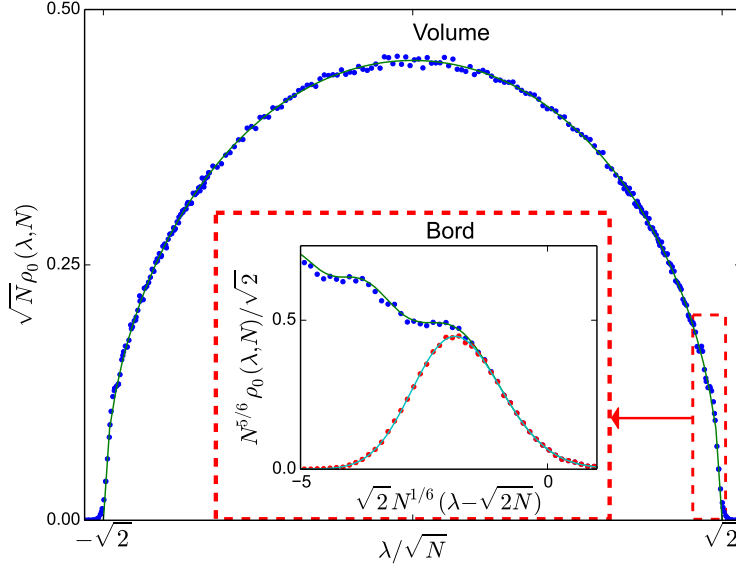


FIGURE 6.1: Comparaison entre les deux régimes de (6.16) (lignes) et des résultats numériques obtenus en simulant $2 \cdot 10^5$ matrices gaussiennes unitaires de taille 150×150 (points). Lorsque qu'on considère des valeurs propres d'ordre \sqrt{N} , on observe le régime de volume (6.12). En venant zoomer proche de la valeur propre maximale $|\lambda - \sqrt{2N}| = \mathcal{O}(N^{-1/6})$, on observe pour la densité le noyau d'Airy au même point (6.15) (points bleus). Dans ce régime, on peut également observer la distribution de la valeur propre maximale (points rouges) donnée par la distribution de Tracy-Widom (6.21).

où $\kappa = 2^{2/3}$ et où Ai est la fonction d'Airy (12). De la même façon, on peut analyser le comportement proche de la valeur propre minimale (qui se trouve à des fluctuations près en $-\sqrt{2N}$) et l'on trouve un comportement symétrique. Ainsi, pour résumer, pour $\lambda > 0$ et dans la limite N grand, la densité moyenne possède les deux régimes

$$\rho_0(\lambda, N) \sim \begin{cases} \frac{1}{\sqrt{N}} \rho_{0,B} \left(\frac{\lambda}{\sqrt{N}} \right), & \lambda = \mathcal{O}(\sqrt{N}) \text{ \& } \lambda < \sqrt{2N}, \\ \sqrt{2} N^{-5/6} \rho_{0,E} \left((\lambda - \sqrt{2N}) \sqrt{2} N^{1/6} \right), & |\lambda - \sqrt{2N}| = \mathcal{O}(N^{-1/6}). \end{cases} \quad (6.16)$$

Tous ces comportements sont repris sur la figure 6.1. Remarquons que les deux régimes de l'équation (6.16) se raccordent bien. Il faut pour cela analyser le régime de volume lorsque l'on se rapproche du bord $\lambda = \sqrt{2N} - \varepsilon$ avec $\varepsilon \rightarrow 0$ ainsi que le régime de bord dans la limite où on s'en éloigne $\lambda = \sqrt{2N} - \varepsilon$ avec $\varepsilon \rightarrow \infty$

$$\rho_0(\sqrt{2N} - \varepsilon, N) = \begin{cases} \frac{1}{\sqrt{N}} \rho_{0,B} \left(\frac{\sqrt{2N} - \varepsilon}{\sqrt{N}} \right) & \xrightarrow{\varepsilon \rightarrow 0} \\ \sqrt{2} N^{-5/6} \rho_{0,E} \left(-\varepsilon \sqrt{2} N^{1/6} \right) & \xrightarrow{\varepsilon \rightarrow \infty} \end{cases} \sim \frac{2^{3/4}}{\pi} N^{-3/4} \sqrt{\varepsilon}. \quad (6.17)$$

où l'on a utilisé le comportement asymptotique de $\rho_{0,E}$ que l'on peut obtenir à partir des expressions explicites (6.15) [64]

$$\rho_{0,E}(x) \underset{x \rightarrow -\infty}{\sim} \frac{1}{\pi} \sqrt{-x}. \quad (6.18)$$

6.4 Loi de Tracy-Widom

Dans la limite $N \rightarrow \infty$, la cumulative de la valeur propre maximale λ_{\max} d'une matrice gaussienne $N \times N$ converge vers la loi de Tracy et Widom [154] lorsque l'on étudie le régime de bord avec une échelle appropriée (le même que le régime de bord (6.16))

$$\lim_{N \rightarrow \infty} \mathbb{P}[\lambda_{\max} < \sqrt{2N} + \frac{x}{\sqrt{2N^{1/6}}}] = \mathcal{F}_\beta(x) , \quad (6.19)$$

où les distributions \mathcal{F}_β s'expriment en fonction de l'unique solution $q(x)$, dite solution de Hastings-McLeod, de l'équation de Painlevé de type II

$$\begin{cases} q''(x) = 2q^3(x) + xq(x) , \\ q(x) \sim \text{Ai}(x) \text{ pour } x \rightarrow \infty . \end{cases} \quad (6.20)$$

On peut montrer [154] que ces distributions s'écrivent explicitement pour les cas gaussiens $\beta = 1, 2$ et 4

$$\mathcal{F}_2(x) = \exp \left(- \int_x^\infty du (u - x) q^2(u) \right) , \quad (6.21)$$

$$\mathcal{F}_1(x) = \exp \left(- \frac{1}{2} \int_x^\infty du q(u) \right) \sqrt{\mathcal{F}_2(x)} , \quad (6.22)$$

$$\mathcal{F}_4(x) = \cosh \left(- \frac{1}{2} \int_x^\infty du q(u) \right) \sqrt{\mathcal{F}_2(x)} . \quad (6.23)$$

Une formule analogue valable pour tous β n'est pour l'instant pas connue. Les fonctions de Hastings-McLeod q et de Tracy-Widom \mathcal{F}_β sont tabulées dans la référence [133]. La figure 6.1 montre une comparaison avec des données numériques pour la densité de probabilité de λ_{\max} , qui est la dérivée de l'équation (6.19). Remarquons que la queue à x grand de $\rho_{0,B}(x)$ [64] correspond bien avec la queue droite de la distribution de Tracy-Widom [65, 109, 117]. Cette loi de Tracy-Widom est un des piliers de la théorie des statistiques d'extrêmes de variables aléatoires fortement corrélées car elle apparaît dans de nombreux problèmes. En effet, on s'attend à ce que la distribution de la valeur propre maximale (ainsi que le régime de bord (6.15)) soit la distribution de Tracy-Widom si la densité dans le régime de volume s'annule proche du bord comme une racine [2, 42]. Ce sera ainsi le cas pour les matrices de Wishart que nous définirons dans le chapitre 9 lorsque l'on analyse la valeur propre maximale (ou la valeur propre minimale dans le régime de *soft edge*) [90, 92]. Notons néanmoins qu'il existe des modèles de matrices aléatoires pour lesquels la densité dans le régime de volume s'annule autrement qu'en racine proche du bord, dans ce cas, les fluctuations de la valeur propre maximale sont distribuées suivant une loi dite de Tracy-Widom d'ordre plus élevé [31]. Le cadre d'application de la loi de Tracy-Widom va même bien au-delà des matrices aléatoires, citons ici quelques exemples non exhaustifs [140] :

- Baik *et al.* [5] ont montré que la distribution de Tracy-Widom décrit la statistique de la plus longue sous-séquence croissante d'une permutation aléatoire. On se donne une permutation σ des n premiers entiers $\{1, 2, \dots, n\}$. On cherche alors dans cette permutation les sous-séquences croissantes et plus particulièrement la longueur l_n de la sous-séquence croissante la plus longue. Par exemple, la permutation pour $\sigma_5 = \{2, 1, 4, 3, 5\}$ possède par exemple comme sous-séquences croissantes : $\{2, 4, 5\}$, $\{1, 3, 5\}$ et $\{1, 5\}$ et la longueur de la plus longue est $l_5 = 3$. Si l'on considère une mesure uniforme sur l'ensemble des $n!$ permutations de $\{1, 2, \dots, n\}$, Baik *et al.* [5] ont montré que

$$\lim_{n \rightarrow \infty} \text{Pr} . [l_n < 2\sqrt{n} + xn^{1/6}] = \mathcal{F}_2(x) . \quad (6.24)$$

• La distribution de Tracy-Widom apparaît également dans le modèle du polymère dirigé que nous avons introduit en introduction (figure 1). Dans ce cas, les fluctuations du chemin de plus basse énergie sont gouvernées par la distribution de Tracy-Widom [6, 90]. Cette distribution apparaît dans les modèles de la classe d'universalité Kardar-Parisi-Zhang comme par exemple le modèle de croissance "Polynuclear Growth Model" [130, 131] ou des modèles d'interfaces [136]. Ces résultats ont été confirmés expérimentalement par Takeuchi et Sano [152]. Dans leur expérience, Takeuchi et Sano [152] mesurent l'interface entre deux phases turbulentes de cristaux liquides et ont montré que ses fluctuations sont données par la distribution de Tracy-Widom dans le régime des grands temps.

• Comme nous l'avons déjà mentionné, cette distribution de Tracy-Widom intervient dans le modèle des ponts et des excursions répulsives introduit dans la section 4.2.2. En effet, l'équation (4.39) se réécrit plus précisément

$$\frac{x_{n,B}(\tau)}{\sqrt{2\tau(1-\tau)}} = \sqrt{2n} + \frac{\chi_2}{\sqrt{2}n^{1/6}} + O(n^{-1/3}), \quad (6.25)$$

où χ_2 est une variable aléatoire distribuée selon la distribution de Tracy-Widom $\Pr. [\chi_2 < x] = \mathcal{F}_2(x)$. Mentionnons le lien avec le processus d'Airy₂ nécessaire pour vérifier à la fois (4.41) et (6.25)

$$\Pr. [\mathcal{A}_2(0) < x] = \mathcal{F}_2(x). \quad (6.26)$$

De plus, dans le cas des excursions répulsives, il a été montré [67] que, dans la limite d'un grand nombre d'excursions, les fluctuations de la plus haute excursion $x_{n,E}$ sont également gouvernées à un temps fixé par la distribution de Tracy-Widom \mathcal{F}_2 tandis que les fluctuations du maximum global sont gouvernées par la distribution de Tracy-Widom \mathcal{F}_1

$$\Pr. \left[\left(\frac{x_{n,E}(\tau)}{\sqrt{2\tau(1-\tau)}} - 2\sqrt{n} \right) 2n^{1/6} \right] = \mathcal{F}_2(x), \quad (6.27)$$

$$\Pr. \left[\max_{0 \leq \tau \leq 1} x_{n,E}(\tau) < \sqrt{2n} \left(1 + \frac{x}{2^{7/3}n^{2/3}} \right) \right] = \mathcal{F}_1(x). \quad (6.28)$$

Ce résultat renforce ainsi le caractère universel de la distribution de Tracy-Widom.

• La statistique de Tracy-Widom apparaît également dans l'état fondamental d'un ensemble de N fermions 1D sans spin confinés dans un potentiel harmonique $V(x) = x^2/2$ [51, 111]. En effet, cet état fondamental Ψ_0 est construit à partir du déterminant de Slater des N premiers états propres à une particule $\psi_i(x) \propto H_n(x)e^{-x^2/2}$ de l'oscillateur harmonique

$$\Psi_0(\vec{x}) = \frac{1}{\sqrt{N!}} \det[\psi_i(x_j)]. \quad (6.29)$$

Il est alors facile d'écrire la distribution jointe des positions des N fermions comme

$$|\Psi_0(\vec{x})|^2 = \frac{1}{Z_N} \prod_{i < j} (x_i - x_j)^2 \exp \left(- \sum_{j=1}^N x_j^2 \right), \quad (6.30)$$

avec Z_N une constante de normalisation. Les positions des N fermions sont donc distribuées suivant la même loi de probabilité jointe que les valeurs propres de l'ensemble gaussien unitaire (6.2). Ainsi, les fluctuations de positions du fermion le plus à droite sont gouvernées par la distribution de Tracy-Widom \mathcal{F}_2 .

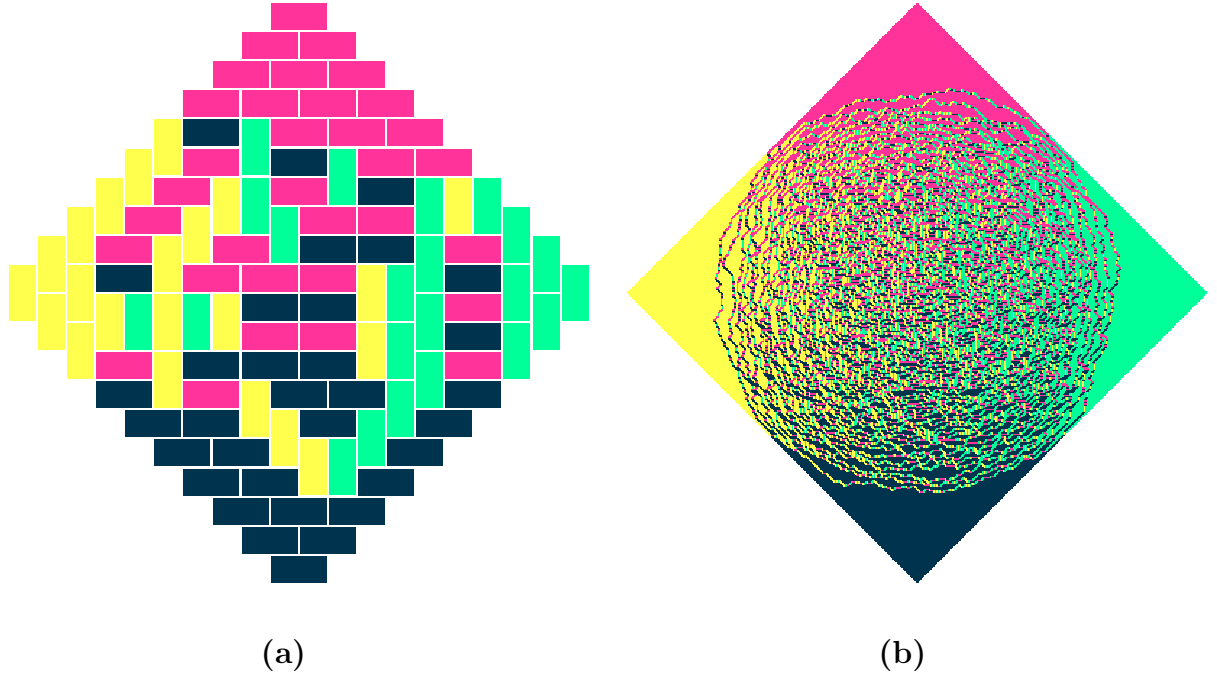


FIGURE 6.2: Exemple de pavages du diamant aztèque par des dominos obtenus en utilisant l'algorithme de glissement de dominos. Nous utilisons 4 couleurs de dominos suivant si celui-ci est horizontal ou vertical et suivant la parité de la case en haut à gauche recouverte par le domino. **(a)** Exemple d'un pavage du diamant de taille $n = 10$ parmi les $2^{n(n+1)/2} \approx 3.6 \times 10^{15}$ pavages possibles. **(b)** Pavage typique du diamant de taille $n = 200$, on observe une zone désordonnée au centre tandis que les bords sont bien rangés. La frontière entre ces deux zones est un cercle dont les fluctuations sont gouvernées par la distribution de Tracy-Widom.

- Mentionnons pour terminer une application à un problème classique de combinatoire : le pavage du diamant aztèque par des dominos [91]. Le diamant aztèque d'ordre n est l'ensemble des cases carrées $[m, m+1] \times [l, l+1]$ où les indices $(m, l) \in \mathbb{Z}^2$ sont à l'intérieur de la région $\{(x, y) \mid |x| + |y| \leq n+1\}$. On cherche alors à paver ce diamant par des dominos de taille 2×1 horizontaux ou verticaux sans qu'ils ne se chevauchent. Nous donnons en figure 6.2 **(a)** un exemple de tel pavage pour le diamant de taille $n = 10$. On peut montrer qu'il existe $2^{n(n+1)/2}$ façons différentes de paver le diamant d'ordre n [52]. Dans le cas où l'on génère ces pavages avec une loi uniforme sur toutes ces possibilités (via l'algorithme de glissement de dominos par exemple [135]), on observe une zone désordonnée au centre du diamant (on parle de *région tempérée*) tandis que les dominos des bords du diamant sont rangés (on parle de *région polaire*) pour les grands diamants comme le montre la figure 6.2 **(b)** pour un pavage typique du diamant de taille 200. Il est possible de montrer que la frontière entre ces deux régions est, dans la limite de n grand, un cercle (le *cercle arctique*) [89] et que les fluctuations de cette frontière sont gouvernées par la distribution de Tracy-Widom [91]. Mentionnons une application de ce modèle à la mécanique statistique d'interfaces élastiques sur laquelle j'ai travaillé durant mon stage de M2 qui a précédé ma thèse et qui a donné lieu à une publication dans Physical review letters [125].

6.5 Transition de phase associée à λ_{\max} : le modèle de May

Nous introduisons dans cette section le modèle de May [112] pour rediscuter les différents régimes dans un langage de thermodynamique, en particulier de transition de phase [108]. Le régime de bord est alors vu comme des corrections à N fini à la transition. Ce langage sera également l'occasion de discuter les fonctions de grandes déviations.

On considère dans ce modèle N espèces ρ_i qui sans interaction sont stables dont on note ρ_i^* les N densités d'équilibres. On suppose les temps d'équilibre des différentes espèces identiques pris à l'unité. En notant les écarts à l'équilibre $x_i(t) = \rho_i(t) - \rho_i^*$, on a

$$\partial_t x_i = -x_i(t). \quad (6.31)$$

On ajoute alors des interactions entre les espèces sous la forme

$$\partial_t x_i = -x_i(t) + \alpha \sum_{j=1}^n M_{i,j} x_j(t), \quad (6.32)$$

où α est une constante de couplage et M une matrice aléatoire d'interaction $N \times N$ que l'on prend gaussienne. Sous quelles conditions le système est-il resté stable ? Pour le savoir, on peut réécrire les N équations (6.32) sous la forme matricielle

$$\partial_t \vec{x} = (\alpha M - I) \vec{x}(t), \quad (6.33)$$

où $\vec{x}(t)$ est le vecteur colonne des écarts à l'équilibres $x_i(t)$ et I est la matrice identité. En notant λ_i les N valeurs propres de la matrice M , le système est stable si

$$\alpha \lambda_i - 1 < 0 \quad \forall i, \quad (6.34)$$

c'est-à-dire si la valeur propre maximale $\lambda_{\max} = \max\{\lambda_i | i \in [1, N]\}$ est bornée

$$\lambda_{\max} < \frac{1}{\alpha}. \quad (6.35)$$

Or, comme nous l'avons vu (6.19), cette valeur propre maximale dans la limite thermodynamique d'un grand nombre N d'espèces tend vers une variable de Tracy-Widom χ_β après renormalisation

$$\lambda_{\max} = \sqrt{2N} + \frac{\chi_\beta}{\sqrt{2}N^{1/6}}. \quad (6.36)$$

Ainsi, dans la limite thermodynamique, on observe une transition de phase pour la probabilité que le système soit stable $P_{\text{stable}}(\alpha, N) = \Pr. [\lambda_{\max} \alpha < 1]$

$$\lim_{N \rightarrow \infty} P_{\text{stable}}(\alpha \sqrt{N}, N) = \begin{cases} 1 & \text{si } \alpha < \sqrt{2}, \\ 0 & \text{si } \alpha > \sqrt{2}. \end{cases} \quad (6.37)$$

Dans ce contexte, le régime de volume correspond au régime thermodynamique. Le régime de bord, auquel correspond la distribution de Tracy-Widom, correspond quant à lui aux corrections à N fini. En utilisant le résultat (6.36), la transition entre la phase stable et la phase instable n'est pas brutale à N fini mais est lisse, donnée par la loi de Tracy-Widom, avec une largeur d'ordre $N^{-1/6}$. Les comportements critiques ont la particularité de gommer les détails et d'être universels. Ainsi, cette analyse renforce donc la conjecture selon laquelle la loi de Tracy-Widom serait associée aux fluctuations critiques de certaines

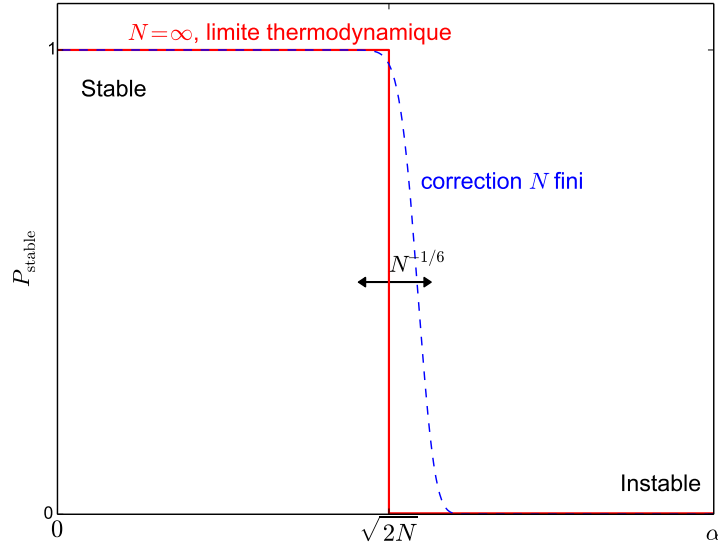


FIGURE 6.3: On observe une transition pour $\alpha = \sqrt{2N}$ entre les phases stable et instable dans la limite thermodynamique. À N fini, la transition n'est pas brutale mais est donnée par la loi de Tracy-Widom avec une largeur d'ordre $N^{-1/6}$.

catégories de transition de phase [140]. Ce comportement est schématisé sur la figure 6.3. Mentionnons que nous n'avons discuté que du comportement typique de la valeur propre maximale. Il est également possible d'étudier ses fluctuations atypiques c'est-à-dire lorsque celle-ci est très loin de $\sqrt{2N}$. On peut montrer [41, 110], en utilisant le gaz de Coulomb (6.6) que les comportements atypiques de la valeur propre maximale s'écrivent pour les ensembles gaussiens

$$\Pr. \left[\frac{\lambda_{\max}}{\sqrt{N}} < x \right] = \begin{cases} e^{-\beta N^2 \phi_-(x)} & \text{si } x < \sqrt{2} \text{ et } |x - \sqrt{2}| \sim O(1), \\ \mathcal{F}_\beta \left(\sqrt{2} N^{2/3} (x - \sqrt{2}) \right) & \text{si } |x - \sqrt{2}| \sim O(N^{-2/3}), \\ 1 - e^{-\beta N \phi_+(x)} & \text{si } x > \sqrt{2} \text{ et } |x - \sqrt{2}| \sim O(1), \end{cases} \quad (6.38)$$

où les fonctions de grandes déviations ϕ_- et ϕ_+ sont indépendantes de β et sont données par

$$\begin{aligned} \phi_-(x) &= \frac{1}{108} \left(36x^2 - x^4 - (15x + x^3)\sqrt{x^2 + 6} + 27 \left(\ln 18 - 2 \ln \left(x + \sqrt{x^2 + 6} \right) \right) \right), \\ \phi_+(x) &= \frac{1}{2} x \sqrt{x^2 - 2} + \ln \left(\frac{x - \sqrt{x^2 - 2}}{\sqrt{2}} \right). \end{aligned} \quad (6.39)$$

En utilisant les comportements (6.38), on peut réécrire

$$\lim_{N \rightarrow \infty} -\frac{1}{N^2 \beta} \log \Pr. \left[\frac{\lambda_{\max}}{\sqrt{N}} < x \right] = \begin{cases} \phi_-(x) & \text{si } x < \sqrt{2}, \\ 0 & \text{si } x > \sqrt{2}, \end{cases} \quad (6.40)$$

qui s'interprète comme l'énergie libre du gaz de Coulomb (6.6). Puisque $\phi_-(x) \sim (\sqrt{2} - x)^3$ pour $x \rightarrow \sqrt{2}$ par valeur inférieure, la dérivée troisième de l'énergie libre est discontinue

au point critique $x = \sqrt{2}$. La transition (6.37) est donc une transition du troisième ordre [108].

Chapitre 7

Statistique à partir de la valeur propre maximale pour l'ensemble gaussien unitaire

Nous allons dans ce chapitre étudier la densité à partir du maximum que nous avons introduite dans l'introduction (5) dans le cas où la collection de variables aléatoires correspond aux valeurs propres d'une matrice gaussienne unitaire et sont donc fortement corrélées. Nous verrons dans la section 7.1 que cette densité est reliée au gap entre les deux plus grandes valeurs propres. En utilisant la méthode des polynômes orthogonaux, nous donnerons ensuite dans la section 7.2 une formule valable à N fini pour la densité à partir du maximum et pour la statistique du premier gap. Nous étudierons alors dans la section 7.3 la limite thermodynamique de grand N de ces formules. Comme pour le cas de la densité que nous avons discuté dans le chapitre précédent, il existe un régime de volume lorsque l'on considère des valeurs propres à une distance $r \sim \sqrt{N}$ du maximum et un régime de bord lorsque l'on considère des valeurs propres proches du maximum d'une distance $r = o(N^{-1/6})$. Nous terminerons ce chapitre en confrontant dans la section 7.4 nos résultats avec des simulations numériques. L'ensemble de ce travail a été publié dans Journal of Statistical Physics [126]. Nous discuterons dans le chapitre suivant d'une application de nos résultats à un modèle de verre de spin en champ moyen.

7.1 Densité à partir de la valeur propre maximale et gap entre les deux plus grandes valeurs propres

De manière analogue à la densité (6.8), on introduit encore une fois la densité d'état à partir de la valeur propre maximale $\rho_{\max}(r, N)$ qui compte pour une matrice gaussienne unitaire de taille $N \times N$ le nombre moyen de valeurs propres à une distance r de la valeur propre maximale (figure 7.1)

$$\rho_{\max}(r, N) = \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq i_{\max}}}^N \langle \delta(\lambda_{\max} - \lambda_i - r) \rangle, \quad (7.1)$$

où i_{\max} est tel que $\lambda_{i_{\max}} = \lambda_{\max}$ est la valeur propre maximale et $\langle \dots \rangle$ correspond à la moyenne suivant le poids (6.2). On a normalisé cette quantité de manière à avoir

$$\int_0^\infty dr \rho_{\max}(r, N) = 1. \quad (7.2)$$

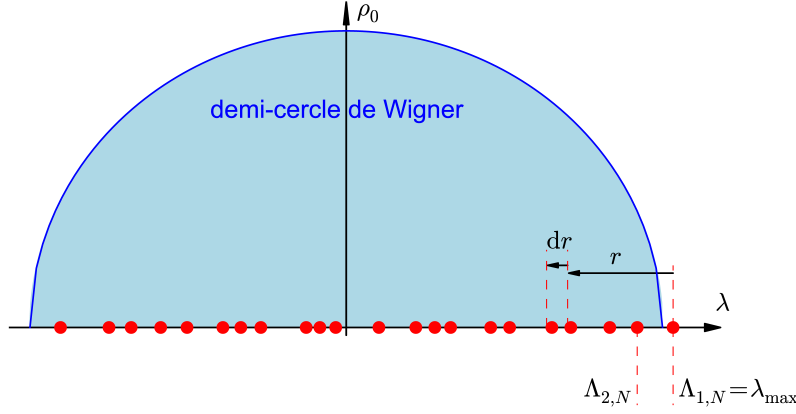


FIGURE 7.1: La densité d'état à partir de la valeur propre maximale $\rho_{\max}(r, N)dr$ compte le nombre moyen de valeurs propres dans l'intervalle $[\lambda_{\max} - r - dr, \lambda_{\max} - r]$. La densité de probabilité du gap entre les deux plus grandes valeurs propres $p_{\text{GAP}}(r, N)$ est donnée par $p_{\text{GAP}}(r, N)dr = \Pr. [\Lambda_{1,N} - \Lambda_{2,N} \in [r, r + dr]]$.

On peut donner une formule plus explicite pour $\rho_{\max}(r, N)$ sous la forme d'une intégrale multiple. En utilisant l'invariance de la densité jointe de probabilité des valeurs propres $P_{\text{joint}}(\lambda_1, \dots, \lambda_N)$ (6.2) par permutation des λ_i , tous les termes de (7.1) sont égaux. Il y a N possibilités pour le choix de la valeur propre maximale dont la valeur y n'est pas fixée et peut donc varier entre $-\infty$ et ∞ tandis que les autres valeurs propres ne peuvent prendre que des valeurs inférieures à cette valeur propre maximale. Une de ces $N - 1$ valeurs propres étant fixée à $y - r$ via la distribution de Dirac. On peut donc réécrire la densité moyenne à partir de la valeur propre maximale $\rho_{\max}(r, N)$ comme

$$\rho_{\max}(r, N) = N \underbrace{\int_{-\infty}^{+\infty} dy}_{N-2 \text{ autres valeurs propres sont inférieures à } \lambda_{\max}} \underbrace{\int_{-\infty}^y d\lambda_1 \cdots \int_{-\infty}^y d\lambda_{N-2}}_{\text{une valeur propre à } r \text{ de } \lambda_{\max}} P_{\text{joint}}(\lambda_1, \dots, \lambda_{N-2}, \underbrace{y-r, y}_{\lambda_{\max}}). \quad (7.3)$$

Une autre quantité intéressante est le premier gap entre les deux plus grandes valeurs propres. Si l'on dénote les valeurs propres ordonnées par $\lambda_{\max} = \Lambda_{1,N} \geq \Lambda_{2,N} \geq \dots \geq \Lambda_{N,N}$ on peut calculer la densité de probabilité $p_{\text{GAP}}(r, N)dr = \Pr. [\Lambda_{1,N} - \Lambda_{2,N} \in [r, r + dr]]$ comme le montre la figure 7.1. De la même façon, une expression exacte pour $p_{\text{GAP}}(r, N)$ peut être écrite facilement via la densité jointe des N valeurs propres (6.2) en intégrant sur les valeurs de $y \in (-\infty, +\infty)$ pour la seconde valeur propre $\Lambda_{2,N}$, tandis que $\Lambda_{1,N} = \lambda_{\max}$ est fixée à $y + r$. Les $(N - 2)$ autres valeurs propres pouvant prendre n'importe quelle valeur plus petite que y . Ainsi

$$p_{\text{GAP}}(r, N) = N(N-1) \underbrace{\int_{-\infty}^{+\infty} dy}_{\text{les } N-2 \text{ autres valeurs propres sont inférieures à } \Lambda_{2,N}} \underbrace{\int_{-\infty}^y d\lambda_1 \cdots \int_{-\infty}^y d\lambda_{N-2}}_{\text{le premier gap est } r} P_{\text{joint}}(\lambda_1, \dots, \lambda_{N-2}, \underbrace{y, y+r}_{\substack{\Lambda_{2,N} \quad \Lambda_{1,N}}}). \quad (7.4)$$

où le facteur $N(N-1)$ dans l'équation (7.4) provient des $N(N-1)$ possibilités pour choisir la paire de valeurs propres $(\Lambda_{1,N}, \Lambda_{2,N})$ parmi les N variables aléatoires. En comparant ce résultat avec l'expression (7.3) et en utilisant l'invariance de la densité jointe des valeurs propres (6.2) par permutation, on obtient la relation

Lien entre la densité à partir du maximum et la statistique du premier gap

$$p_{\text{GAP}}(r, N) = (N - 1)\rho_{\text{max}}(-r, N) . \quad (7.5)$$

Remarquons donc que cela nécessite d'analyser la densité ρ_{max} en dehors de son intervalle naturel de définition $[0, \infty[$, de telle recherche ont déjà été étudiées dans d'autres buts [68]. Remarquons également le rôle différent que joue la variable muette y qui est la plus grande valeur propre dans (7.3) et la seconde dans (7.4).

7.2 Formule à N fini

Pour calculer l'intégrale multiple (7.3), nous allons utiliser le formalisme des polynômes orthogonaux [74]. Nous introduisons donc la famille étagée de polynômes unitaires $\pi_k(\lambda, y)$ qui sont des polynômes de la variable λ dépendant du paramètre y . Ces polynômes sont définis par la relation [118]

$$\begin{cases} \langle \pi_k, \pi_{k'} \rangle = \int_{-\infty}^y d\lambda \pi_k(\lambda, y) \pi_{k'}(\lambda, y) e^{-\lambda^2} = \delta_{k,k'} h_k(y) , \\ \pi_k(\lambda, y) = \lambda^k + \dots \end{cases} \quad (7.6)$$

Quand $y \rightarrow \infty$, ces polynômes orthogonaux dégénèrent en polynômes d'Hermite (20). Pour y fini, il n'existe pas de formule simple pour les calculer. Il est néanmoins possible de calculer les premiers polynômes [118]

$$\begin{aligned} \pi_0(\lambda, y) &= 1 , \\ \pi_1(\lambda, y) &= \lambda + a , \\ \pi_2(\lambda, y) &= \lambda^2 + \lambda \left(\frac{a + y}{1 - 2a(a + y)} - y \right) - 1 + \frac{1}{2 - 4a(a + y)} , \\ \pi_3(\lambda, y) &= \lambda^3 + \left[\lambda^2 2a (8a^2 - 4a^2 y^2 + 8ay - 8ay^3 - 3 - 4y^4) \right. \\ &\quad + 2\lambda (-12a^3 y - 10a^2 - 22a^2 y^2 - 9ay - 10ay^3 + 3) \\ &\quad + a (-16a^2 + 4a^2 y^2 - 20ay + 8ay^3 + 5 - 4y^2 + 4y^4) \Big] \\ &\quad / (8a^3 y + 12a^2 + 16a^2 y^2 + 12ay + 8ay^3 - 4) , \end{aligned} \quad (7.7)$$

avec $a = \frac{e^{-y^2}}{\sqrt{\pi}(1+\text{erf}(y))}$ et $\text{erf}(y)$ la fonction d'erreur (11). En utilisant les polynômes (7.7), on peut aussi calculer les premiers h_k

$$\begin{aligned} h_0 &= e^{-y^2} \frac{1}{2a} , \\ h_1 &= e^{-y^2} \frac{1}{4} \left(\frac{1}{a} - 2a - 2y \right) , \\ h_2 &= e^{-y^2} \frac{2a^3 y + 3a^2 + 4a^2 y^2 + 3ay + 2ay^3 - 1}{4a(2a^2 + 2ay - 1)} , \\ h_3 &= e^{-y^2} \left(-32a^4 + 4a^4 y^2 - 60a^3 y + 16a^3 y^3 + 29a^2 - 20a^2 y^2 + 20a^2 y^4 + 30ay \right. \\ &\quad \left. + 8ay^3 + 8ay^5 - 6 \right) / (16a(2a^3 y + 3a^2 + 4a^2 y^2 + 3ay + 2ay^3 - 1)) . \end{aligned} \quad (7.8)$$

Les expressions des polynômes $\pi_k(\lambda, y)$ deviennent de plus en plus compliquées quand k devient grand. Pour analyser ces polynômes quand k tend vers l'infini, il est intéressant d'introduire la relation de récurrence à trois termes qu'ils satisfont [147]. En effet,

$\lambda \pi_k(\lambda, y)$ étant un polynôme unitaire de degré $k + 1$, on peut le décomposer sur la base étagée $\{\pi_k(\lambda, y)\}_{i=0,1,\dots,k+1}$. Cependant, les produits scalaires $\langle \lambda \pi_k, \pi_i \rangle = \langle \pi_k, \lambda \pi_i \rangle$ sont nuls dès que $i \leq k - 2$. En posant

$$S_k(y) h_k(y) = \langle \lambda \pi_k, \pi_k \rangle, \quad (7.9)$$

$$R_k(y) h_{k-1}(y) = \langle \lambda \pi_k, \pi_{k-1} \rangle, \quad (7.10)$$

on peut donc écrire

$$\lambda \pi_k(\lambda, y) = \pi_{k+1}(\lambda, y) + S_k(y) \pi_k(\lambda, y) + R_k(y) \pi_{k-1}(\lambda, y). \quad (7.11)$$

En réinjectant l'équation (7.10) dans l'équation de récurrence (7.11) prise au rang $k - 1$, on trouve directement

$$R_k(y) = \frac{h_k(y)}{h_{k-1}(y)}. \quad (7.12)$$

Une intégration par parties permet de calculer le produit scalaire $\langle \lambda \pi_k, \pi_k \rangle$

$$\langle \lambda \pi_k, \pi_k \rangle = \int_{-\infty}^y dy \lambda \pi_k^2(\lambda, y) e^{-\lambda^2} \quad (7.13)$$

$$= \int_{-\infty}^y dy \pi_k(\lambda, y) \partial_\lambda \pi_k(\lambda, y) e^{-\lambda^2} - \left[\frac{1}{2} \pi_k^2(\lambda, y) e^{-\lambda^2} \right]_{\lambda=-\infty}^{\lambda=y} \quad (7.14)$$

$$= -\frac{1}{2} \pi_k^2(y, y) e^{-y^2} = -\frac{1}{2} \partial_y h_k. \quad (7.15)$$

car $\partial_y \pi_k(\lambda, y)$ est un polynôme en λ de degré $k - 1$ puisque $\pi_k(\lambda, y)$ est un polynôme unitaire en λ (7.6), ce qui implique $\langle \partial_y \pi_k, \pi_k \rangle = 0$. Ce résultat permet d'exprimer le coefficient S_k

$$S_k(y) = -\partial_y \ln h_k(y) / 2. \quad (7.16)$$

La densité $\rho_{\max}(r, N)$ dans (7.3) peut être exprimée en fonction de ces polynômes orthogonaux $\pi_k(\lambda, y)$. Il suffit de remplacer le déterminant de Vandermonde présent dans la probabilité jointe (6.2) par un déterminant construit à partir des polynômes orthogonaux $\pi_k(\lambda, y)$

$$\prod_{i < j} |\lambda_i - \lambda_j| = |\det(\lambda_i^j)| = |\det(\pi_j(\lambda_i, y))|. \quad (7.17)$$

On écrit alors le déterminant en utilisant son développement de Laplace et on utilise la condition d'orthogonalité (7.6) pour calculer les intégrales sur $\lambda_1, \dots, \lambda_{N-2}$ dans (7.3). Grâce à ces manipulations standard [65, 113], on trouve que $\rho_{\max}(r, N)$ peut être écrit comme [157, 158]

$$\rho_{\max}(r, N) = \frac{N(N-2)!}{Z_N} \int_{-\infty}^{\infty} dy \prod_{k=0}^{N-1} h_k(y) \begin{vmatrix} K_N(y-r, y-r) & K_N(y-r, y) \\ K_N(y, y-r) & K_N(y, y) \end{vmatrix}, \quad (7.18)$$

où $K_N(\lambda, \lambda')$ est le noyau (*Kernel*) associé aux polynômes orthogonaux (7.6) défini par

$$K_N(\lambda, \lambda') = \sum_{k=0}^{N-1} \frac{1}{h_k(y)} \pi_k(\lambda, y) \pi_k(\lambda', y) e^{-\frac{\lambda^2 + \lambda'^2}{2}} = \sum_{k=0}^{N-1} \psi_k(\lambda, y) \psi_k(\lambda', y), \quad (7.19)$$

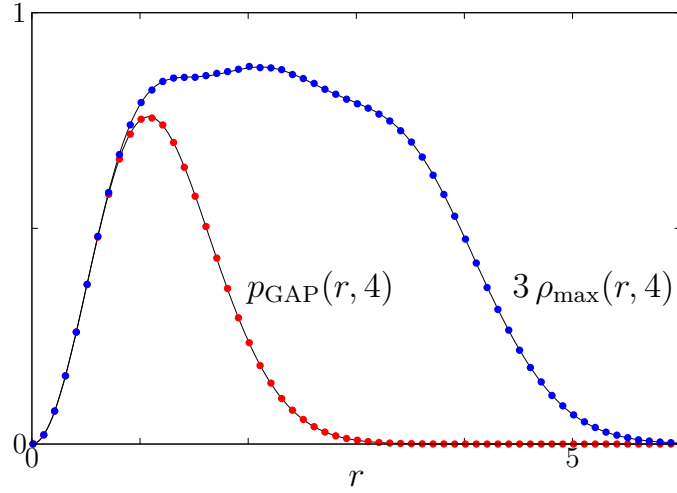


FIGURE 7.2: Résultats numériques obtenus en simulant $2 \cdot 10^7$ matrices gaussiennes unitaires de taille 4×4 (points) comparés aux résultats analytiques obtenus via l'équation (7.18) (trait noir).

où nous avons introduit les fonctions d'ondes normalisées

$$\psi_k(\lambda, y) = \frac{1}{\sqrt{h_k(y)}} \pi_k(\lambda, y) e^{-\frac{\lambda^2}{2}}. \quad (7.20)$$

En utilisant la formule de Christoffel-Darboux, on peut réécrire le noyau $K_N(\lambda, \lambda')$

$$K_N(\lambda, \lambda') = \sqrt{R_N(y)} \frac{\psi_N(\lambda, y) \psi_{N-1}(\lambda', y) - \psi_{N-1}(\lambda, y) \psi_N(\lambda', y)}{\lambda - \lambda'}. \quad (7.21)$$

Remarquons que, même si nous ne l'avons pas noté pour alléger la notation, le noyau dépend du paramètre y . Notons également que même si la densité $\rho_{\max}(r, N)$ est naturellement définie pour $r > 0$ par l'équation (6.7), elle admet néanmoins une extension naturelle pour $r < 0$ en utilisant l'expression (7.18). La formule (7.18) peut être évaluée pour des petites valeurs de N en utilisant les expressions des équations (7.7) et (7.8). La figure 7.2 compare les résultats numériques obtenus en simulant un grand nombre de matrices gaussiennes unitaires de taille 4×4 avec la formule (7.18). Pour réécrire la formule (7.18), nous dérivons une identité sur la distribution cumulée de la plus grande valeur propre $F_N(y) = \Pr.[\lambda_{\max} \leq y]$

$$F_N(y) = \int_{-\infty}^y d\lambda_1 \cdots \int_{-\infty}^y d\lambda_N P_{\text{joint}}(\lambda_1, \dots, \lambda_N). \quad (7.22)$$

En remplaçant le déterminant de Vandermonde par un déterminant construit avec les polynômes orthogonaux (7.6) et en utilisant la formule de Cauchy-Binet, on trouve [65, 113, 118]

$$F_N(y) = \frac{N!}{Z_N} \prod_{j=0}^{N-1} h_j(y). \quad (7.23)$$

En dérivant cette équation par rapport à y , on obtient

$$\partial_y F_N(y) = \frac{N!}{Z_N} \sum_{k=0}^{N-1} \left(\partial_y h_k(y) \prod_{j=0, j \neq k}^{N-1} h_j(y) \right) = \frac{N!}{Z_N} \left(\prod_{j=0}^{N-1} h_j(y) \right) \sum_{k=0}^{N-1} \frac{\partial_y h_k(y)}{h_k(y)}. \quad (7.24)$$

D'un autre côté, en utilisant le résultat de l'intégration par parties (7.13), on peut donc réécrire cette identité

$$\partial_y \ln F_N(y) = \sum_{k=0}^{N-1} \frac{\pi_k^2(y, y)}{h_k(y)} e^{-y^2} = K_N(y, y) . \quad (7.25)$$

En utilisant les résultats (7.23) et (7.25), la densité à partir du maximum $\rho_{\max}(r, N)$ (7.18) peut se réécrire

— Densité à partir du maximum pour des matrices gaussiennes unitaires —

$$\rho_{\max}(r, N) = \frac{1}{N-1} \int_{-\infty}^{\infty} dy \left[F'_N(y) K_N(y-r, y-r) - F_N(y) K_N^2(y, y-r) \right] , \quad (7.26)$$

où l'on a utilisé la propriété de symétrie $K_N(\lambda, \lambda') = K_N(\lambda', \lambda)$. Remarquons que le deuxième terme de (7.26) provient des corrélations entre la plus grande valeur propre λ_{\max} en y et la valeur propre en $y-r$. On peut aussi vérifier la normalisation de $\rho_{\max}(r, N)$ en utilisant les identités

$$\int_0^{+\infty} K_N(y-r, y-r) dr = \int_{-\infty}^y K_N(x, x) dx = N , \quad (7.27)$$

$$\int_0^{+\infty} K_N^2(y, y-r) dr = K_N(y, y) = \partial_y \ln F_N(y) , \quad (7.28)$$

où on a utilisé les identités (7.25). Remarquons que, dans la limite de N grand, le poids du premier terme de (7.26) est d'ordre $\mathcal{O}(1)$ tandis que le second est d'ordre sous dominant $\mathcal{O}(N^{-1})$.

7.3 Limite à grand N

Nous allons désormais faire une analyse à grand N de la densité à partir du maximum $\rho_{\max}(r, N)$ et, en utilisant la relation (7.5), de la densité de probabilité du premier gap $p_{\text{GAP}}(r, N)$. Nous avons vu dans l'introduction qu'il existait deux régimes différents pour la densité comme le montre l'équation (6.16) et la figure 6.1. L'un est le régime dit de volume, pour des valeurs propres λ_i d'ordre \sqrt{N} , tandis que l'autre régime est dit de bord quand les valeurs propres sont proches du bord de demi-cercle de Wigner, $\lambda - \sqrt{2N} = \mathcal{O}(N^{-1/6})$. Ainsi en analysant le comportement à grand N de $\rho_{\max}(r, N)$, on s'attend à observer également ces deux régimes, comme le montre la figure 7.3, quand $r \propto \sqrt{N}$ et quand $r = \mathcal{O}(N^{-1/6})$.

7.3.1 Régime $r \propto \sqrt{N}$: régime de volume

Dans ce régime, il est facile de voir que le deuxième terme de (7.26) est sous-dominant par rapport au premier. Nous avons déjà vu que le poids total de celui-ci est seulement d'ordre $\mathcal{O}(N^{-1})$ tandis que le premier terme a un poids d'ordre $\mathcal{O}(1)$. En effet, le second terme provient des corrélations entre la valeur propre maximale λ_{\max} , située en y , avec la valeur propre en $y-r$ qui est dans ce régime d'ordre \sqrt{N} . Pour analyser le premier terme de (7.26) on peut remplacer la densité de probabilité de λ_{\max} , $F'_N(y)$ par un pic de Dirac $\delta(y - \sqrt{2N})$. En effet, le noyau $K_N(y-r, y-r)$ avec $(y-r) \propto \sqrt{N}$ n'est pas sensible aux fluctuations typiques d'ordre $\mathcal{O}(N^{-1/6})$ de λ_{\max} (6.36) autour $y = \sqrt{2N}$. En notant $r = \hat{r}\sqrt{N}$, on a donc

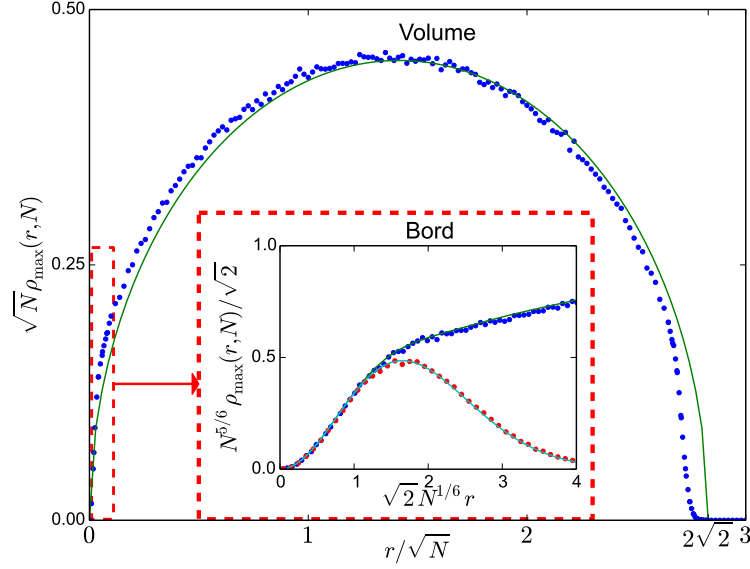


FIGURE 7.3: Visualisation des deux régimes (lignes) et des résultats numériques obtenus en simulant $2 \cdot 10^5$ matrices gaussiennes unitaires de taille 150×150 (points). Lorsque l'on considère des valeurs propres à une distance d'ordre $r = \sqrt{N}$ du maximum, on observe le régime de volume explicité en section 7.3.1. En venant zoomer proche de la valeur propre maximale, c'est-à-dire dans le régime où $r = \mathcal{O}(N^{-1/6})$, on observe le régime de bord discuté en section 7.3.2 pour la densité proche du maximum (points rouges) et le régime typique du premier gap (points bleus).

— Régime de volume : loi du demi-cercle translaté —

$$\begin{aligned} \rho_{\max}(\hat{r}\sqrt{N}, N) &\sim \frac{1}{N} K_N(\sqrt{2N} - \hat{r}\sqrt{N}, \sqrt{2N} - \hat{r}\sqrt{N}) \underset{N \rightarrow \infty}{\sim} \frac{1}{\sqrt{N}} \tilde{\rho}_B \left(\hat{r} = \frac{r}{\sqrt{N}} \right), \\ \tilde{\rho}_B(\hat{r}) &= \rho_W(\sqrt{2} - \hat{r}) = \frac{1}{\pi} \sqrt{\hat{r}(2\sqrt{2} - \hat{r})}, \end{aligned} \quad (7.29)$$

Ce résultat (7.29) peut être facilement interprété, la densité de valeurs propres à une distance r de $\lambda_{\max} \sim \sqrt{2N}$ dans (7.1) avec $(r - \lambda_{\max}) \propto \sqrt{N}$ est insensible aux fluctuations de λ_{\max} qui sont d'ordre $\mathcal{O}(N^{-1/6})$. Par conséquent, $\rho_{\max}(r, N)$ est tout simplement le demi-cercle de Wigner décalé (7.29).

7.3.2 Régime $r = \mathcal{O}(N^{-1/6})$: régime de bord

Dans ce régime, il est nécessaire d'analyser le noyau $K_N(y - r, y - r')$ avec à la fois $y \sim \sqrt{2N}$ et r, r' d'ordre $N^{-1/6}$. Cela correspond à une analyse de double échelle des relations de récurrence (7.11). Nadal et Majumdar [118] ont utilisé ce système de polynômes orthogonaux (7.6) pour calculer la distribution cumulative de la valeur propre maximale λ_{\max} (7.23) qui peut être exprimée en termes des racines des normes $h_k(y)$ uniquement. Pour retrouver la distribution de Tracy-Widom, il faut effectuer l'analyse des $h_k(y)$ pour des k d'ordre N , avec à la fois $N \rightarrow \infty$ et $y - \sqrt{2N} = \mathcal{O}(N^{-1/6})$ [118]. Cette limite de double échelle des $h_k(y)$ a été par la suite généralisée à d'autres modèles de matrices [1] comme par exemple pour les matrices de Wishart que nous étudierons dans le chapitre

9. Ici, en plus des $h_k(y)$, le calcul de la densité à partir du maximum $\rho_{\max}(r, N)$ (7.18) nécessite d'analyser le noyau $K_N(y - r, y - r')$ dans cette limite de double échelle. Dans ce but, nous analysons la relation de récurrence à trois termes (7.11) que nous réécrivons comme une relation de récurrence satisfaite par les fonctions d'onde $\psi_N(\lambda, y)$ définie par l'équation (7.20)

$$\lambda \psi_N(\lambda, y) = \sqrt{R_{N+1}(y)} \psi_{N+1}(\lambda, y) + S_N(y) \psi_N(\lambda, y) + \sqrt{R_N(y)} \psi_{N-1}(\lambda, y) . \quad (7.30)$$

On définit alors la variable x comme

$$y - \sqrt{2N} = \frac{x}{\sqrt{2}} N^{-\frac{1}{6}} . \quad (7.31)$$

Dans cette limite de double échelle, les coefficients $R_N(y)$ et $S_N(y)$ prennent la forme [118]

$$\begin{aligned} R_N(y) &= \frac{N}{2} \left(1 - N^{-\frac{2}{3}} q^2(x) + \mathcal{O}(N^{-1}) \right) , \\ S_N(y) &= -\frac{N^{-\frac{1}{6}}}{\sqrt{2}} q^2(x) + \mathcal{O}(N^{-\frac{1}{2}}) . \end{aligned} \quad (7.32)$$

Dans la limite $y \gg \sqrt{2N}$, on peut remplacer y par $+\infty$ dans l'intégrale qui définit les polynômes $\pi_k(\lambda, y)$ (7.6), ces polynômes dégènèrent ainsi en polynômes d'Hermite (20) [118]

$$\psi_N(\lambda, y) = \frac{1}{\pi^{1/4} 2^{N/2} \sqrt{N!}} H_N(\lambda) e^{-\frac{\lambda^2}{2}} + \mathcal{O}(e^{-y^2}) . \quad (7.33)$$

Nous voulons maintenant évaluer la fonction d'onde $\psi_N(y - r, y)$ quand y atteint la valeur $\sqrt{2N}$ par dessus, à partir du développement à grand y (7.33) . En posant $y = \sqrt{2N} + xN^{-1/6}/\sqrt{2}$ et en utilisant la formule de Plancherel-Rotach pour les polynômes d'Hermite [147] nous trouvons que $\psi_N(y - r, y)$ a pour forme limite

$$\psi_N(y - r, y) \underset{N \rightarrow \infty}{\sim} 2^{1/4} N^{-1/12} \text{Ai}(x - \sqrt{2} N^{1/6} r) + \mathcal{O}(e^{-y^2}) , \quad (7.34)$$

qui est donc une fonction des deux variables d'échelle

$$\begin{cases} x = \sqrt{2} N^{1/6} (y - \sqrt{2N}) , \\ \tilde{r} = \sqrt{2} N^{1/6} r . \end{cases} \quad (7.35)$$

Compte tenu de ce résultat, il est naturel de rechercher une solution des relations de récurrence (7.30) sous la forme

$$\psi_N(y - r, y) = \frac{2^{1/4}}{\sqrt{\pi}} N^\nu G\left(\sqrt{2} N^{1/6} r, \sqrt{2} N^{1/6} (y - \sqrt{2N})\right) , \quad (7.36)$$

où l'exposant ν et la fonction G sont encore à déterminer et où l'amplitude $2^{1/4}/\sqrt{\pi}$ est choisie ici pour plus de commodité. En faisant correspondre les deux formules (7.36) et (7.34) on trouve

$$\nu = -\frac{1}{12} , \quad G(\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} \sqrt{\pi} \text{Ai}(x - \tilde{r}) . \quad (7.37)$$

L'étape suivante est alors d'insérer l'ansatz (7.36) dans la relation de récurrence (7.30) pour ensuite effectuer un développement en puissance de N dans la limite de double

échelle, correspondant au régime (7.35). Pour ce faire, nous aurons également besoin des développements de $x_N = \sqrt{2}N^{1/6}(y - \sqrt{N})$ et de $r_N = \sqrt{2}N^{1/6}r$

$$\begin{cases} x_{N+1} = x_N - N^{-1/3} + \mathcal{O}(N^{-1}) \\ r_{N+1} = r_N + \mathcal{O}(N^{-1}) . \end{cases} \quad (7.38)$$

En utilisant ces deux relations, le développement de la fonction d'onde s'écrit

$$\psi_{N\pm 1}(y - r, y) = \frac{2^{1/4}}{\sqrt{\pi}N^{1/12}} \left(G(\tilde{r}, x) \mp N^{-\frac{1}{3}} \partial_x G(\tilde{r}, x) + N^{-\frac{2}{3}} \partial_x^2 G(\tilde{r}, x) + \mathcal{O}(N^{-1}) \right) . \quad (7.39)$$

En injectant dans la relation de récurrence (7.30) l'ansatz (7.36) ainsi que la fonction d'onde (7.39) et en utilisant les résultats (7.32), on trouve que, au premier ordre, la fonction $G(\tilde{r}, x)$ est solution de l'équation de Schrödinger :

$$-\partial_x^2 G(\tilde{r}, x) + [x + 2q^2(x)]G(\tilde{r}, x) = \tilde{r} G(\tilde{r}, x) . \quad (7.40)$$

Il est possible de montrer que la fonction $\tilde{f}(\tilde{r}, x) = 2^{-1/6}G(\tilde{r}, x)$ vérifiant les conditions déduites des équations (7.40) et (7.37)

$$\begin{aligned} \partial_x^2 \tilde{f}(\tilde{r}, x) - [x + 2q^2(x)]\tilde{f}(\tilde{r}, x) &= -\tilde{r}\tilde{f}(\tilde{r}, x) , \\ \tilde{f}(\tilde{r}, x) &\underset{x \rightarrow \infty}{\sim} 2^{-1/6} \sqrt{\pi} \text{Ai}(x - \tilde{r}) , \end{aligned} \quad (7.41)$$

est le premier membre d'une paire de Lax associée à une équation de Painlevé XXXIV [126]. En utilisant le développement de la fonction d'onde (7.39), la forme limite du noyau (7.21) dans la limite de double échelle (7.35) s'écrit

$$K_N(y - r, y - r') = \frac{1}{\pi} \frac{G(\tilde{r}, x) \partial_x G(\tilde{r}', x) - \partial_x G(\tilde{r}, x) G(\tilde{r}', x)}{(\tilde{r}' - \tilde{r})/(\sqrt{2}N^{1/6})} . \quad (7.42)$$

Après un long calcul [126], en partie reproduit dans l'appendice F, la densité à partir du maximum dans le régime de bord $r = \tilde{r}N^{-1/6}/\sqrt{2}$ s'écrit finalement

—— Densité à partir du maximum dans le régime de bord ——

$$\rho_{\max}(r, N) = N^{-5/6} \sqrt{2} \tilde{\rho}_E(\sqrt{2}N^{1/6}r) , \quad (7.43)$$

$$\tilde{\rho}_E(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx , \quad (7.44)$$

dont on peut analyser les comportements asymptotiques

—— Comportement asymptotique du régime de bord ——

$$\tilde{\rho}_E(\tilde{r}) \sim \begin{cases} \frac{1}{2} \tilde{r}^2 + a_4 \tilde{r}^4 + o(\tilde{r}^4) , & \tilde{r} \rightarrow 0 \\ \frac{\sqrt{\tilde{r}}}{\pi} , & \tilde{r} \rightarrow \infty , \end{cases} \quad (7.45)$$

où l'amplitude a_4 est donnée par une intégrale [126]. Le comportement asymptotique à grand \tilde{r} (7.45) garantit une parfaite adéquation entre le régime de volume décrit par la

loi du demi-cercle de Wigner translaté (7.29) proche du bord $r \rightarrow 0$ et le régime de bord $\tilde{\rho}_E$ (7.43) en direction du volume $r \rightarrow \infty$

$$\rho_{\max}(r, N) = \begin{cases} \frac{1}{\sqrt{N}} \tilde{\rho}_B\left(\frac{r}{\sqrt{N}}\right) & \xrightarrow{r \rightarrow 0} \\ N^{-5/6} \sqrt{2} \tilde{\rho}_E(\sqrt{2} N^{1/6} r) & \xrightarrow{r \rightarrow \infty} \end{cases} \sim \frac{2^{3/4} \sqrt{\tilde{r}}}{n^{3/4} \pi} . \quad (7.46)$$

Le comportement quadratique de $\tilde{\rho}_E(\tilde{r})$ à petit \tilde{r} peut être interprété. En effet, ce comportement est lié à la probabilité que les deux plus grandes valeurs propres deviennent extrêmement proches (7.5) qui, en utilisant la densité de probabilité conjointe des valeurs propres (6.2), doit bien s'annuler quadratiquement pour l'ensemble des matrices gaussiennes unitaires. Remarquons que cet argument ne permet pas de déterminer le facteur $1/2$ de (7.45) qui est un résultat non trivial de notre calcul. En utilisant la relation (7.5), l'analyse de double échelle effectuée pour calculer $\tilde{\rho}_E(\tilde{r})$ (7.43) peut être directement adaptée pour obtenir une formule pour la distribution du gap typique $p_{\text{GAP}}(r, N)$ entre les deux plus grandes valeurs propres, c'est-à-dire lorsque $r = \mathcal{O}(N^{-1/6})$ par opposition aux fluctuations atypiques associées aux grandes déviations. On trouve directement, dans la limite de grand N

— Statistique typique du premier gap —

$$p_{\text{GAP}}(r, N) = \sqrt{2} N^{1/6} \tilde{p}_{\text{typ}}\left(r \sqrt{2} N^{1/6}\right) , \quad (7.47)$$

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(-\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx . \quad (7.48)$$

Le changement de signe $\tilde{r} \rightarrow -\tilde{r}$ a des conséquences sur les comportements asymptotiques [126]

— Comportement asymptotique du premier gap —

$$\tilde{p}_{\text{typ}}(\tilde{r}) \sim \begin{cases} \frac{1}{2} \tilde{r}^2 + a_4 \tilde{r}^4 + o(\tilde{r}^4) , & \tilde{r} \rightarrow 0 \\ \frac{2^{-91/48} e^{\zeta'(-1)}}{\sqrt{\pi}} e^{-\frac{4}{3} \tilde{r}^{3/2} + \frac{8}{3} \sqrt{2} \tilde{r}^{3/4}} \tilde{r}^{-\frac{21}{32}} \left(1 - \frac{1405 \sqrt{2}}{1536} \tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2}) \right) , & \tilde{r} \rightarrow \infty , \end{cases} \quad (7.49)$$

avec $\zeta'(-1) = -0.165\dots$ la dérivée de la fonction zêta de Riemann (19) au point -1 . Le comportement à grand \tilde{r} peut être interprété, à l'ordre dominant, en supposant que les deux plus grandes valeurs propres $\Lambda_{1,N} = \lambda_{\max}$ et $\Lambda_{2,N} = \Lambda_{1,N} - r$ sont indépendantes (car l'écart entre les deux est grand). En utilisant cette hypothèse, on a alors

$$\begin{aligned} p_{\text{GAP}}(r, N) &= \int_{-\infty}^{\infty} \text{Pr} . [\Lambda_{2,N} = \lambda, \lambda_{\max} = \lambda + r] d\lambda \\ &\underset{r \rightarrow \infty}{\sim} \int_{-\infty}^{\infty} \text{Pr} . [\Lambda_{2,N} = \lambda] \text{Pr} . [\lambda_{\max} = \lambda + r] d\lambda \\ &\underset{r \rightarrow \infty}{\sim} \text{Pr} . [\lambda_{\max} = r] . \end{aligned} \quad (7.50)$$

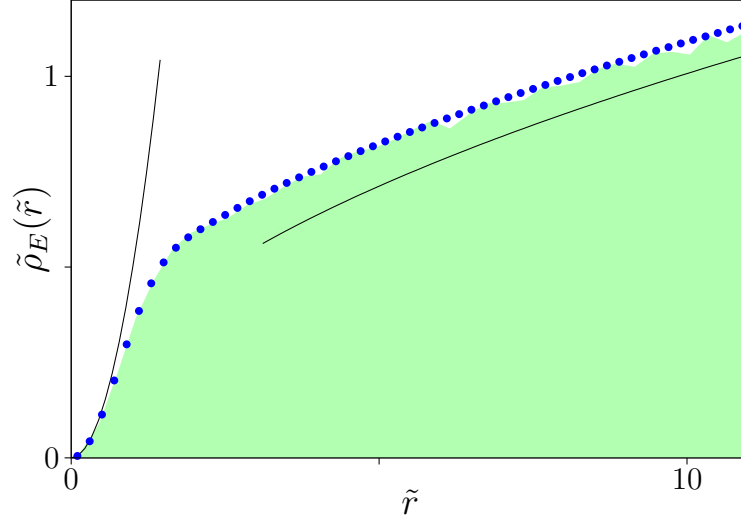


FIGURE 7.4: Points bleus : évaluation numérique de notre formule (7.43) pour $\tilde{\rho}_E(\tilde{r})$, remplissage vert : résultats numériques obtenus en simulant 2×10^5 matrices gaussiennes unitaires de taille 1000×1000 , ligne noire solide : comportements asymptotiques donnés dans l'équation (7.45).

Ainsi, à l'ordre dominant, le comportement asymptotique de $\tilde{p}_{\text{typ}}(\tilde{r})$ à grand \tilde{r} est le même que la queue droite de Tracy-Widom : $\ln \tilde{p}_{\text{typ}}(\tilde{r}) \sim -\frac{4}{3}\tilde{r}^{3/2} + o(\tilde{r}^{3/2})$. Il est également intéressant de remarquer que notre résultat est différent du *Wigner surmise* qui décrit les écarts entre les valeurs propres moyennés sur le régime de volume. Dans le régime de volume, l'écart moyen entre les valeurs propres est d'ordre $\sim N^{-1/2}$ et est donné par la formule approchée dite de *Wigner surmise* (pour $\beta = 2$) [164] (exacte pour $n = 2$)

$$P(s) = \frac{32}{\pi^2} s^2 \exp\left(-\frac{4}{\pi} s^2\right). \quad (7.51)$$

Il existe également une formule dans la limite de grand n , dite de Gaudin-Mehta [113], exprimant cette quantité en fonction d'une solution de l'équation de Painlevé V. Notre résultat analyse le premier gap, dans le régime de bord. L'écart typique entre les valeurs propres y est beaucoup plus grand, d'ordre $\sim N^{-1/6}$ et le comportement à grand argument est différent d'une gaussienne et est donné par une exponentielle étirée.

7.4 Simulations numériques

7.4.1 Simulation rapide des matrices gaussiennes

Il est possible de simuler les matrices gaussiennes en utilisant la définition (6.1). Néanmoins [44] ont trouvé un modèle de matrices symétriques réelles tridiagonales qui possède la même loi jointe des valeurs propres que celle des ensembles gaussiens (6.4). La partie

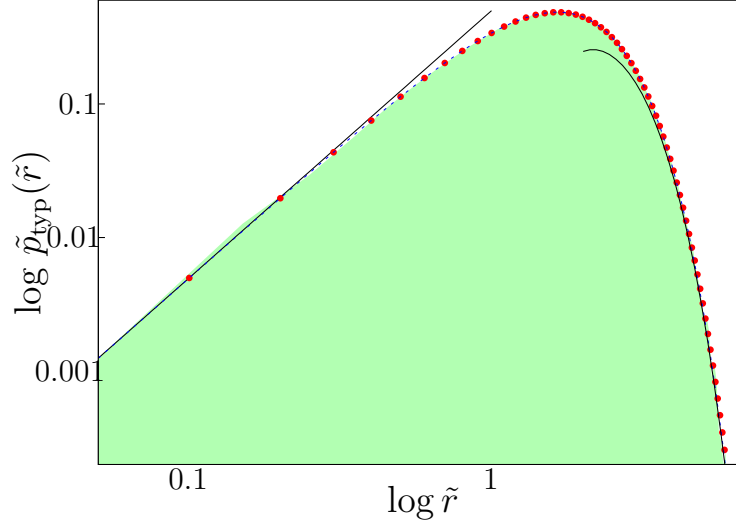


FIGURE 7.5: Points rouges : évaluation de notre formule (7.47) pour $\tilde{p}_{\text{typ}}(\tilde{r})$ en échelle log-log, courbe noire en pointillés : résultat numérique provenant de [166], remplissage vert : résultats numériques obtenus en simulant 2×10^5 matrices gaussiennes unitaires de taille 1000×1000 , ligne noire solide : comportements asymptotiques donnés dans l'équation (7.49).

triangulaire supérieure de ces matrices est donnée par

$$\begin{pmatrix} \mathcal{N}[0, \frac{1}{\sqrt{2}}] & \frac{1}{2}\chi_{\beta(N-1)} & 0 & \cdots & 0 \\ & \mathcal{N}[0, \frac{1}{\sqrt{2}}] & \frac{1}{2}\chi_{\beta(N-2)} & \ddots & \vdots \\ & & \mathcal{N}[0, \frac{1}{\sqrt{2}}] & \ddots & 0 \\ & & & \ddots & \frac{1}{2}\chi_{\beta} \\ & & & & \mathcal{N}[0, \frac{1}{\sqrt{2}}] \end{pmatrix} \quad (7.52)$$

avec χ_d la racine d'une loi χ^2 à d degrés de liberté, c'est-à-dire la loi de la somme de d lois normales aux carrées. Cette méthode permet de générer les valeurs propres pour le cas gaussien beaucoup plus rapidement car la matrice est désormais tridiagonale symétrique réelle. Elle permet également de générer des variables aléatoires corrélées suivant (6.4) pour tout β et pas uniquement pour les cas $\beta = 1, 2$ et 4 .

7.4.2 Évaluation numérique des fonctions d'échelle

Dans cette section, nous fournissons une évaluation numérique des fonctions $\tilde{\rho}_E(\tilde{r})$ et $\tilde{p}_{\text{typ}}(\tilde{r})$ données respectivement dans les équations (7.43) et (7.47). Pour évaluer numériquement ces formules, nous avons besoin de calculer numériquement la solution \tilde{f} . Nous avons constaté que la meilleure façon de les évaluer est de résoudre numériquement (avec Mathematica) le système d'équations différentielles satisfaites par $\tilde{f}(\tilde{r}, x)$ (7.41), \tilde{r} étant un paramètre fixé avec q la solution de Hasting-McLeod de l'équation de Painlevé II (6.20) dont les données numériques sont tirées des tables [133]. Une fois que $\tilde{f}(\tilde{r}, x)$ a été évaluée numériquement, on peut calculer les distributions $\tilde{\rho}_E(\tilde{r})$ et $\tilde{p}_{\text{typ}}(\tilde{r})$ en évaluant numériquement les intégrales des équations (7.43) et (7.47). La figure 7.4 montre la fonction d'échelle $\tilde{\rho}_E(\tilde{r})$, évaluée numériquement (les points bleus), en fonction de \tilde{r} . Les lignes noires indiquent le comportement au premier ordre à petit \tilde{r} , $\tilde{\rho}_E(\tilde{r}) \sim \tilde{r}^2/2$

(7.45) ainsi que le premier ordre à grand \tilde{r} , $\tilde{\rho}_E(\tilde{r}) \sim \sqrt{\tilde{r}}/\pi$. Une estimation numérique de $\tilde{\rho}_E(\tilde{r}) - \sqrt{\tilde{r}}/\pi$ indique que cette quantité décroît comme $\tilde{r}^{-1/2}$ pour les grands \tilde{r} . Enfin, le remplissage vert sur la figure 7.4 montre une estimation numérique de $\tilde{\rho}_E(\tilde{r})$ obtenue en simulant 2×10^5 matrices hermitiennes aléatoires indépendantes de taille 1000×1000 . L'accord avec l'évaluation numérique de notre formule (7.43) est assez bon, ce qui fournit un contrôle de cette formule. Il convient toutefois de noter que nous n'avons pas tenté d'évaluer la précision de notre procédure numérique. Nous avons également évalué numériquement notre formule de $\tilde{p}_{\text{typ}}(\tilde{r})$, donnée en équation (7.47). Les valeurs numériques sont représentées par des points rouges sur la figure 7.5. Nous avons comparé ces valeurs à l'évaluation numérique de $\tilde{p}_{\text{typ}}(\tilde{r})$ faite par Witte *et al.* [166] (la courbe en pointillé noir sur la figure 7.5). Cette comparaison montre un bon accord entre l'évaluation numérique de notre formule de l'équation (7.47) et les valeurs tabulées [166]. Dans la figure 7.5, le remplissage vert représente des données obtenues en simulant 2×10^5 matrices gaussiennes unitaires indépendantes de taille 1000×1000 . Enfin, sur la figure 7.5, nous avons également tracé les comportements asymptotiques (7.49) que nous avons obtenus pour les petits et grands \tilde{r} (ligne noire solide). Une analyse plus précise pour les petits et grands \tilde{r} des données [166] montre en fait un très bon accord avec nos asymptotiques précis (7.49).

Chapitre 8

Généralisations et Applications

Nous allons discuter dans ce chapitre de quelques ouvertures et applications de la densité à partir du maximum pour les matrices aléatoires gaussiennes calculée dans le chapitre précédent dans le cas des matrices unitaires $\beta = 2$. Même s'il est difficile d'étendre nos résultats aux cas β quelconque, nous allons dans la section 8.1 donner des arguments permettant de deviner au moins les régimes d'échelles et les comportements asymptotiques. Nous montrerons ensuite dans la section 8.2 que la densité à partir des extrêmes apparaît naturellement dans un problème de minimisation d'un hamiltonien contraint. Cet exemple mène alors au modèle de verre de spin de Sherrington-Kirkpatrick que nous développons dans la section 8.3. Pour terminer ce chapitre, nous verrons dans la section 8.4 que les résultats du chapitre précédent permettent de connaître la densité à partir du maximum ainsi que la distribution du premier gap lorsque la valeur propre maximale est fixée. Ces compléments sont traités dans l'article sur ArXiv [127] ainsi que dans un papier à paraître [124] en collaboration avec Y.V.Fyodorov.

8.1 Ouvertures aux ensembles gaussiens β

Même s'il est difficile d'étendre cette étude pour tous les ensembles gaussiens β , il est possible de donner en utilisant des arguments heuristiques quelques comportements que l'on peut confronter aux données numériques obtenues en générant des matrices tridiagonales introduit dans la section 7.4.1 [44]. On s'attend, pour tous β , à observer encore une fois un régime de volume pour $r \sim \sqrt{N}$ et un régime de bord pour $r = O(N^{-1/6})$

— Régimes de la densité à partir du maximum pour les ensembles gaussiens —

$$\rho_{\max}(r, N) = \begin{cases} \frac{1}{\sqrt{N}} \tilde{\rho}_B\left(\frac{r}{\sqrt{N}}\right) & r \sim \sqrt{N}, \\ N^{-5/6} \sqrt{2} \tilde{\rho}_E(\sqrt{2} N^{1/6} r) & r = O(N^{-1/6}). \end{cases} \quad (8.1)$$

La fonction d'échelle de volume $\tilde{\rho}_B(\hat{r})$ est, en utilisant les mêmes arguments que précédemment, le demi-cercle de Wigner translaté (7.29) indépendamment de β . La fonction d'échelle de bord $\tilde{\rho}_E(\tilde{r})$ dépend quant à elle de l'indice de Dyson β et est difficile à calculer. Néanmoins, on peut prédire les comportements asymptotiques de cette fonction d'échelle. Dans la limite de petit \tilde{r} , la contribution provient uniquement de la valeur propre la plus proche du maximum c'est-à-dire la seconde valeur propre. Dans ce cas, le terme dominant provient du déterminant de Vandermonde de l'équation (6.4), on attend donc le comportement $\tilde{\rho}_E(\tilde{r}) \sim a_\beta \tilde{r}^\beta$ pour l'ensemble gaussien β . Il n'y a aucun argument simple donnant

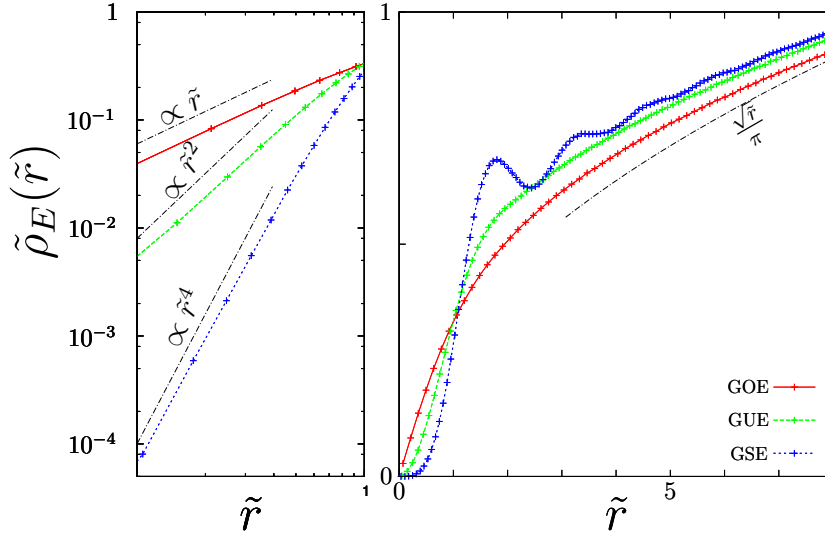


FIGURE 8.1: Tracé de $\tilde{\rho}_E(\tilde{r})$ pour trois valeurs de β ($\beta = 1$ en rouge, $\beta = 2$ en vert et $\beta = 4$ en bleu). Les données sont obtenues en simulant 2.10^7 matrices de taille 100×100 pour chacun des cas. **Figure de gauche** : les données sont tracées dans une échelle log-log. Les comportements à petit \tilde{r} sont conformes à la limite asymptotique (8.2). **Figure de droite** : les comportements de $\tilde{\rho}_E(\tilde{r})$ à grand \tilde{r} sont bien conformes avec le raccordement avec la loi du demi-cercle de Wigner translaté (8.2).

la valeur de la constante a_β (on sait néanmoins que $a_2 = 1/2$ à partir de notre analyse du cas unitaire). D'autre part, si l'on suppose un raccordement en douceur entre la région de bord et la région décrite par la loi du demi-cercle de Wigner décalé (7.29), le comportement à grand \tilde{r} de $\tilde{\rho}_E(\tilde{r})$ est, à l'ordre dominant, $\tilde{\rho}_B(\tilde{r}) \sim \sqrt{\tilde{r}}/\pi$ pour correspondre avec le comportement impliqué par (7.29). Ainsi on attend les comportements asymptotiques suivants pour l'ensemble gaussien β

Comportement asymptotique du régime de bord pour les ensembles gaussiens

$$\tilde{\rho}_B(\tilde{r}) \sim \begin{cases} a_\beta \tilde{r}^\beta + o(\tilde{r}^\beta), & \tilde{r} \rightarrow 0 \\ \frac{\sqrt{\tilde{r}}}{\pi} + o(\tilde{r}^{1/2}), & \tilde{r} \rightarrow \infty. \end{cases} \quad (8.2)$$

On montre dans la figure 8.1 la fonction d'échelle $\tilde{\rho}_E(\tilde{r})$ pour les valeurs naturelles de β des ensembles gaussiens : $\beta = 1$ pour l'ensemble orthogonal **GOE**, $\beta = 2$ pour l'ensemble unitaire **GUE**, $\beta = 4$ pour l'ensemble symplectique **GSE**. Comme on peut le voir, les comportements asymptotiques dictés par l'équation (8.2) sont bien vérifiés à la fois à petit \tilde{r} (figure de gauche) et à grand \tilde{r} (figure de droite). Dans chaque cas, nous avons simulé 2.10^7 matrices indépendantes de taille 100×100 . Mentionnons qu'il apparaît dans la courbe $\tilde{\rho}_E(\tilde{r})$ des oscillations pour $\beta = 4$ alors que les courbes pour $\beta = 1$ et 2 sont beaucoup plus lisses. De la même façon, il est possible de donner les comportements asymptotiques de la distribution du premier gap. Le régime typique du gap étant le régime de bord, on s'attend donc à pouvoir écrire comme pour le cas unitaire (7.47) une fonction d'échelle

Statistique typique du premier gap pour les ensembles gaussiens

$$p_{\text{GAP}}(r, N) = \sqrt{2}N^{1/6} \tilde{p}_{\text{typ}}\left(r\sqrt{2}N^{1/6}\right). \quad (8.3)$$

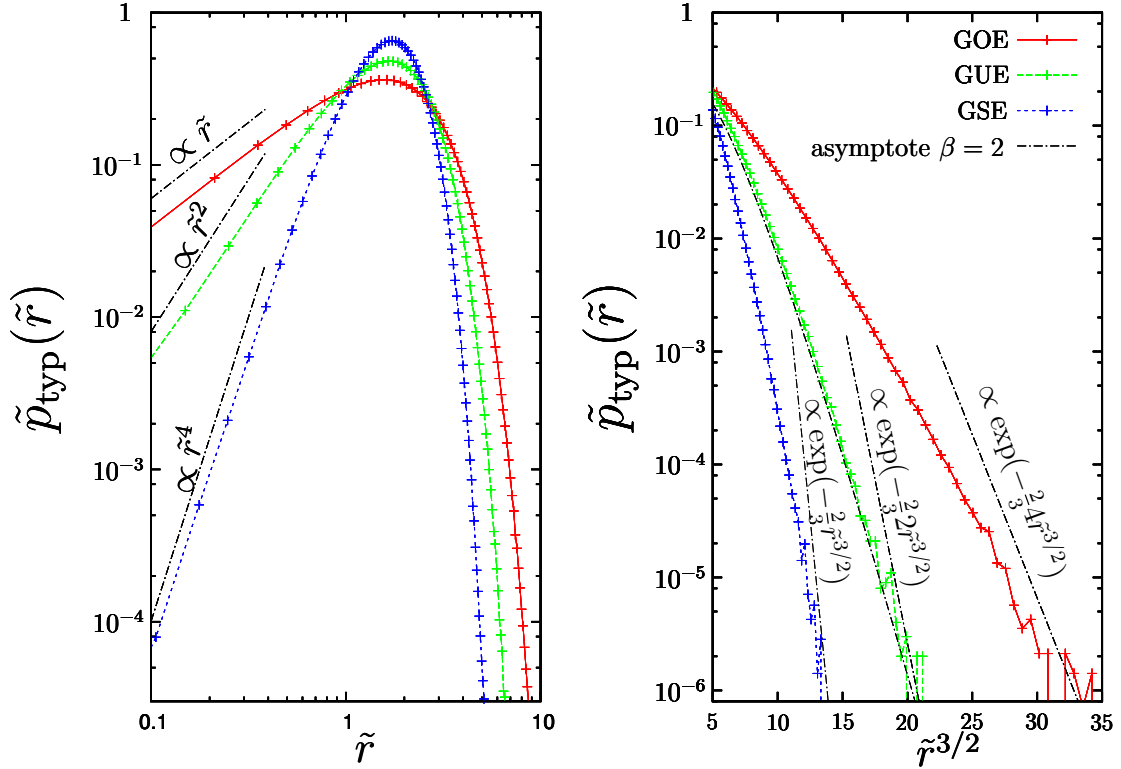


FIGURE 8.2: Tracé de $\tilde{p}_{\text{typ}}(\tilde{r})$ pour trois valeurs de β ($\beta = 1$ en rouge, $\beta = 2$ en vert et $\beta = 4$ en bleu). Les données sont obtenues en simulant 2.10^7 matrices de taille 100×100 pour chacun des cas. **Figure de gauche :** les données sont tracées dans une échelle log-log. Les comportements à petit \tilde{r} sont identiques à ceux de $\tilde{\rho}_E(\tilde{r})$. **Figure de droite :** les données sont tracées dans une échelle logarithmique en fonction de $\tilde{r}^{3/2}$. Les comportements à grand \tilde{r} sont conformes la queue droite de la loi de Tracy-Widom (8.4). Nous avons également indiqué le comportement asymptotique pour $\beta = 2$ (7.49).

Le comportement à petit \tilde{r} , en utilisant la correspondance entre le premier gap et la densité à partir du maximum (7.5), est donné par le même comportement $\tilde{p}_{\text{typ}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta$ (8.2). Le comportement à grand \tilde{r} peut être obtenu, à l'ordre dominant, en supposant que les deux plus grandes valeurs propres $\Lambda_{1,N} = \lambda_{\max}$ et $\Lambda_{2,N} = \Lambda_{1,N} - r$ sont indépendantes lorsque l'écart entre les deux est grand, via l'argument (7.50). Ainsi, à l'ordre dominant, le comportement asymptotique de $\tilde{p}_{\text{typ}}(\tilde{r})$ à grand \tilde{r} est le même que la queue droite de Tracy-Widom

Comportement asymptotique du premier gap pour les ensembles gaussiens

$$\ln \tilde{p}_{\text{typ}}(\tilde{r}) \sim -\frac{2\beta}{3}\tilde{r}^{3/2} + o(\tilde{r}^{3/2}), \quad \tilde{r} \rightarrow +\infty. \quad (8.4)$$

La figure 8.2 montre la fonction d'échelle $\tilde{p}_{\text{typ}}(\tilde{r})$ pour $\beta = 1, 2$ et 4 . La figure de gauche montre le comportement à petit \tilde{r} en échelle log-log tandis que la figure de droite montre le comportement à grand \tilde{r} . Ces deux comportements asymptotiques sont bien en accord avec nos hypothèses. Notons enfin que, en dehors du cas $\beta = 2$, les fonctions d'échelle $\tilde{\rho}_E(\tilde{r})$ et $\tilde{p}_{\text{typ}}(\tilde{r})$ complète, les calculs de la constante a_β ainsi que les corrections asymptotiques à petit et grand arguments ne sont pour l'instant pas connus.

8.2 Application à la minimisation d'une forme quadratique

Ce paragraphe développe une application de la densité à partir du maximum pour les matrices aléatoires. En effet, comme nous allons le voir, cette densité apparaît naturellement dans un problème de minimisation. Ce problème nous permettra également d'introduire le modèle de verre de spin de Sherrington-Kirkpatrick dans la partie suivante.

On définit l'hamiltonien

$$H[\vec{s}] = -\frac{1}{2} \sum_{i,j=1}^N J_{i,j} s_i s_j, \quad (8.5)$$

où \mathbf{J} est une matrice de l'ensemble gaussien β ($\beta = 1, 2$ étant les choix les plus naturels). Nous choisissons dans cette partie ainsi que dans la suivante d'utiliser la convention utilisée par Cugliandolo et Dean [32] où les éléments de la matrice \mathbf{J} sont des gaussiennes centrées de variance d'ordre $1/N$. En utilisant ce choix, le modèle converge sans scaling vers le régime de volume dans la limite thermodynamique. On cherche alors à minimiser cet hamiltonien (8.5) sur la N -sphère c'est-à-dire lorsque l'ensemble des N spins \vec{s} vérifie la contrainte

$$\vec{s}^2 = \sum_{i=1}^N s_i^2 = N. \quad (8.6)$$

Pour prendre en compte cette contrainte (8.6), on introduit le multiplicateur de Lagrange z et la quantité à minimiser, par rapport à \vec{s} et z , est donc

$$\tilde{H}[\vec{s}, z] = -\frac{1}{2} \sum_{i,j=1}^N J_{i,j} s_i s_j + z (\vec{s}^2 - N). \quad (8.7)$$

On aboutit alors à

$$\min_{\vec{s}, z} \tilde{H}[\vec{s}, z] = \tilde{H}[\vec{s}_{\max}, z_{\max}] = -N \frac{\lambda_{\max}}{2}, \quad (8.8)$$

avec \vec{s}_{\max} et z_{\max} donnés par

$$\begin{cases} \mathbf{J} \vec{s}_{\max} = \lambda_{\max} \vec{s}_{\max}, \\ z_{\max} = \frac{\lambda_{\max}}{2}. \end{cases} \quad (8.9)$$

Le spectre de la matrice hessienne $\delta^2 \tilde{H} / \delta s_i \delta s_j$ prise en ce minimum \vec{s}_{\max}, z_{\max} est alors donné par

$$\text{Sp} \left[\frac{\delta^2 \tilde{H}}{\delta s_i \delta s_j} \Big|_{\vec{s}_{\max}, z_{\max}} \right] = \{0, \lambda_{\max} - \lambda_1, \lambda_{\max} - \lambda_2, \dots, \lambda_{\max} - \lambda_N\}. \quad (8.10)$$

Ainsi, la densité à partir du maximum $\rho_{\text{DOS}}(r, N)$ définie dans l'équation (7.1) est la moyenne des valeurs propres de cette matrice hessienne prise au minimum, en excluant la valeur propre 0. Ainsi, et comme nous allons le développer dans la section suivante, la densité à partir du maximum joue un rôle important dans la dynamique des systèmes dont l'hamiltonien est (8.5) contraint par (8.6) qui est le modèle de verre de spin complètement connecté de Sherrington-Kirkpatrick [32, 98].

8.3 Applications à un modèle de verre de spin en champ moyen

L'hamiltonien (8.5) couplé à la contrainte (8.6) constituent le modèle de verre de spin de Sherrington-Kirkpatrick sphérique [32]. Pour étudier la dynamique de ce système, il est naturel de travailler dans la base de vecteurs propres de la matrice J . On note alors $s_\lambda(t)$ la projection de la configuration des spins dans cette base. On considèrera par la suite une condition initiale uniforme. La dynamique de ce système est alors donnée par l'équation de Langevin

$$\begin{aligned}\partial_t s_\lambda(t) &= (\lambda - z(t))s_\lambda(t) + h_\lambda(t) + \xi_\lambda(t), \\ s_\lambda(t=0) &= 1,\end{aligned}\tag{8.11}$$

où $z(t)$ est le paramètre de Lagrange associé à la condition de normalisation (8.6), $h_\lambda(t)$ est un champ magnétique extérieur infinitésimal permettant de calculer la fonction de réponse et $\xi_\lambda(t)$ est un bruit blanc gaussien de température T

$$\overline{\xi_\lambda(t)} = 0, \quad \overline{\xi_\lambda(t)\xi_\mu(t')} = 2T\delta_{\lambda,\mu}\delta(t-t'),\tag{8.12}$$

où \overline{X} correspond à la moyenne sur le bruit thermique tandis que nous garderons la notation $\langle X \rangle$ pour la moyenne sur le désordre c'est-à-dire sur les matrices J (6.4). Pour étudier la dynamique hors d'équilibre de ce système, on peut considérer plusieurs grandeurs en lien avec la densité à partir du maximum ou le premier gap :

- la fonction de corrélation du système qui mesure le recouvrement entre les spins à des temps différents

$$\overline{C(t,t')} = \frac{1}{N} \sum_{i=1}^N \overline{\langle s_i(t)s_i(t') \rangle} = \frac{1}{N} \sum_{\lambda} \overline{\langle s_\lambda(t)s_\lambda(t') \rangle}.\tag{8.13}$$

- la fonction de réponse du système à un champ magnétique extérieur $t > t'$

$$\overline{R(t,t')} = \sum_{i=1}^N \overline{\left. \frac{\delta \langle s_i(t) \rangle}{\delta h_i(t')} \right|_{h=0}} = \sum_{\lambda} \overline{\left. \frac{\delta \langle s_\lambda(t) \rangle}{\delta h_\lambda(t')} \right|_{h=0}}.\tag{8.14}$$

On peut en effet montrer qu'à température nulle ces deux quantités s'écrivent [124]

$$\overline{C(t,t')} = \frac{g_N\left(\frac{t+t'}{2}\right)}{\sqrt{g_N(t)g_N(t')}},\tag{8.15}$$

$$\begin{aligned}\overline{R(t,t')} &= \sqrt{\frac{g_N(t')}{g_N(t)}} \frac{1}{N} \sum_{\lambda \neq \lambda_{\max}} e^{(\lambda - \lambda_{\max})(t-t')} \\ &\quad + \sqrt{\frac{g_N(t')}{g_N(t)}} \frac{1}{N} \left[1 - \frac{\sum_{\lambda} e^{2\lambda t + (\lambda - \lambda_{\max})(t-t')}}{\sum_{\lambda} e^{2\lambda t}} \right], t > t',\end{aligned}\tag{8.16}$$

$$g_N(t) = \frac{1}{N} \sum_{\lambda \neq \lambda_{\max}} e^{2(\lambda - \lambda_{\max})t}.\tag{8.17}$$

En effet, en utilisant la définition de la densité à partir du maximum (7.1), la fonction $\langle g_N(t) \rangle$ est donnée par la transformée de Laplace de cette densité

Application de la densité à partir du maximum au modèle de Sherrington-Kirkpatrick

$$\langle g_N(t) \rangle = \frac{1}{N} \left(1 + (N-1) \int_0^\infty dr e^{-2rt} \rho_{\max}(r, N) \right). \quad (8.18)$$

Il est alors intéressant d'étudier l'influence de nos différents résultats, en particulier des comportements asymptotiques (8.2) sur la dynamique de ce système. Notons que ce contexte fait apparaître des questions d'automoyennage non triviales, il est en effet a priori nécessaire de s'intéresser à la moyenne $\langle g_N(t) \rangle$ pour avoir la relation (8.18). On pourra trouver de plus amples détails dans notre travail en collaboration avec Y. Fyodorov [124]. Nous retrouvons par exemple dans le régime $1 \ll t \ll N^{2/3}$ et $1 \ll t' \ll N^{2/3}$, qui correspond au régime de volume (7.29), le résultat de Cugliandolo et Dean [32]

$$\overline{R(t, t')} \sim \left(\frac{t}{t'} \right)^{3/4} \bar{h} \left(\frac{t - t'}{2} \right) \quad (8.19)$$

$$\bar{h}(t) = \lim_{N \rightarrow \infty} \int_0^\infty dr e^{-2rt} \rho_{\max}(r, N) = \frac{e^{-4t}}{2t} I_1(4t). \quad (8.20)$$

En particulier dans la limite $t - t' \gg 1$ (toujours dans le régime $1 \ll t \ll N^{2/3}$ et $1 \ll t' \ll N^{2/3}$), on trouve

$$R(t, t') \sim \frac{1}{4\sqrt{2\pi}} \left(\frac{t}{t'} \right)^{3/4} \frac{1}{(t - t')^{3/2}}, \quad (8.21)$$

qui est en accord avec la théorie de *local scale invariance* [81]. Dans le cas où les deux temps t et t' sont d'ordre $N^{2/3}$, tout en gardant l'écart $t - t' \sim O(N^{2/3})$, on trouve que le régime de quasi-équilibre [98] est directement donné à partir de la densité proche du maximum dans le régime de bord

$$\overline{R(t, t')} = \frac{1}{N} f_R \left(\frac{t - t'}{N^{2/3}} \right) \quad (8.22)$$

$$f_R(x) = \int_0^\infty d\tilde{r} \tilde{\rho}_E(\tilde{r}) e^{-\frac{x\tilde{r}}{\sqrt{2}}}. \quad (8.23)$$

En particulier, les régimes asymptotiques (8.2) impliquent que, pour $\beta = 1$

$$f_R(x) \propto \begin{cases} x^{-3/2}, & x \ll 1 \\ x^{-2}, & x \gg 1. \end{cases} \quad (8.24)$$

Le comportement à petit x (qui provient du comportement de $\tilde{\rho}_E(\tilde{r})$ à grand \tilde{r}) se raccorde bien avec le régime précédent (8.21) tandis que le régime à grand x correspond quant à lui à un résultat nouveau.

8.4 Densité à partir du maximum et premier gap à λ_{\max} fixée

Nous n'avons étudié pour l'instant que la densité à partir du maximum $\rho_{\text{DOS}}(r, N)$ et la densité de probabilité du premier gap $p_{\text{GAP}}(r, N)$ moyennées sur toutes les valeurs possibles de la valeur propre maximale λ_{\max} (7.3) et (7.4). Il peut également être intéressant d'étudier ces deux quantités lorsque l'on contraint la valeur propre maximale à valoir une certaine valeur y donnée $\rho_{\text{DOS}}(r|y, N)$ et $p_{\text{GAP}}(r|y, N)$. Ce genre de contrainte

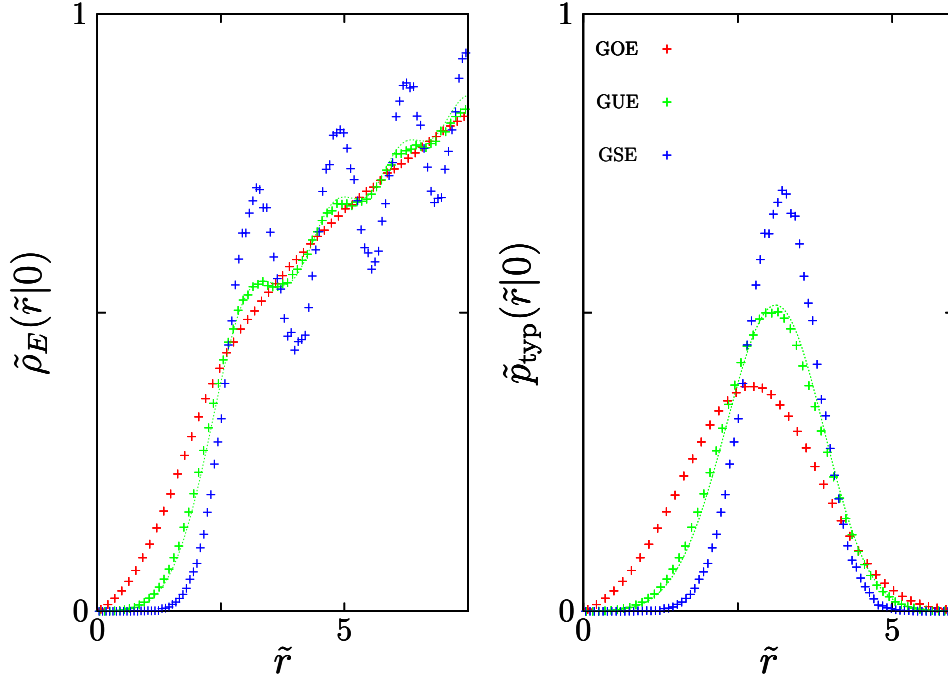


FIGURE 8.3: Tracé de $\tilde{\rho}_{\text{edge}}(\tilde{r}|0)$ (**gauche**) et $\tilde{p}_{\text{typ}}(\tilde{r}|0)$ (**droite**) pour trois valeurs de β ($\beta = 1$ en rouge, $\beta = 2$ en vert et $\beta = 4$ en bleu). Les données numériques (points) sont obtenues en simulant $2 \cdot 10^7$ matrices de taille 200×200 et en ne gardant que les cas où $|\lambda_{\max} - \sqrt{2N}| < 0.1$. Nous donnons également les résultats exacts obtenus pour $\beta = 2$ (8.27) et (8.28) (ligne verte).

se retrouve naturellement en particulier avec le modèle des matrices de Markov aléatoires, matrices dont les éléments sont des variables aléatoires sur $[0, 1]$ et dont la somme sur chaque colonne est 1, qui ont la propriété d'avoir leur valeur propre maximale exactement égale à 1. Ces quantités contraintes doivent posséder les mêmes régimes d'échelle que les cas non contraints, on s'attend donc à avoir pour n'importe quel $\beta > 0$

— Régime de bords lorsque λ_{\max} est fixée —

$$\rho_{\text{DOS}}(r|y, N) = N^{-5/6} \sqrt{2} \tilde{\rho}_E(\sqrt{2}N^{1/6}r|\sqrt{2}N^{1/6}(y - \sqrt{2N})), \quad (8.25)$$

$$p_{\text{GAP}}(r|y, N) = N^{1/6} \sqrt{2} \tilde{p}_{\text{typ}}(\sqrt{2}N^{1/6}r|\sqrt{2}N^{1/6}(y - \sqrt{2N})). \quad (8.26)$$

Dans le cas $\beta = 2$, nous avons obtenu ces quantités [126] au cours de notre calcul dans l'équation (F.16) de l'appendice F.

— Densité à partir du maximum et premier gap lorsque λ_{\max} est fixée —

$$\tilde{\rho}_E(\tilde{r}|x) = \frac{2^{1/3}}{\pi} \mathcal{F}_2(x) \left[R(x) \left(\left(\tilde{r} + \frac{R(x)}{q^2(x)} \right) \tilde{f}^2(\tilde{r}, x) - 2 \frac{q'(x)}{q(x)} \tilde{f}(\tilde{r}, x) \tilde{g}(\tilde{r}, x) \right. \right. \\ \left. \left. + \left(1 + \frac{q^2(x)}{\tilde{r}} \right) \tilde{g}(\tilde{r}, x)^2 \right) - \frac{q^2(x)}{\tilde{r}^2} \tilde{g}^2(\tilde{r}, x) \right], \quad (8.27)$$

$$\tilde{\rho}_{\text{typ}}(\tilde{r}|x) = \frac{2^{1/3}}{\pi} \mathcal{F}_{2,s} \left[R_s \left(\left(-\tilde{r} + \frac{R_s}{q_s^2} \right) \tilde{f}_s^2 - 2 \frac{q'_s}{q_s} \tilde{f}_s \tilde{g}_s + \left(1 - \frac{q_s^2}{\tilde{r}} \right) \tilde{g}_s^2 \right) \right. \\ \left. - \frac{q_s^2}{\tilde{r}^2} \tilde{g}_s^2 \right], \quad (8.28)$$

où $R(x) = \int_x^\infty q^2(u) du$ et $q(x) \tilde{g}(\tilde{r}, x) = -\tilde{r} \int_x^\infty q(u) \tilde{f}(\tilde{r}, u) du$ et l'indice s correspond à des quantités translatées de \tilde{r} c'est-à-dire $q_s = q(x - \tilde{r})$, $R_s = R(x - \tilde{r})$, $\mathcal{F}_{2,s} = \mathcal{F}_2(x - \tilde{r})$, $\tilde{f}_s = \tilde{f}(-\tilde{r}, x - \tilde{r})$, $\tilde{g}_s = \tilde{g}(-\tilde{r}, x - \tilde{r})$. Remarquons que dans cette formule, obtenue en utilisant le lien entre le premier gap et la densité à partir du maximum (7.5), il est nécessaire d'utiliser les quantités translatées à cause du rôle différent de la variable muette y qui est la plus grande valeur propre dans (7.3) et la seconde dans (7.4). En utilisant l'analyse numérique développée dans la section 7.4.2, il est possible de calculer ces quantités et de les confronter à des résultats obtenus en simulant un grand nombre de matrices. Nous traçons dans la figure 8.3 les résultats obtenus lorsque l'on contraint $\lambda_{\max} = \sqrt{2N}$, c'est-à-dire lorsque $x = 0$, pour les cas $\beta = 1, 2$ et 4 . Ces données sont bien en accord avec les formules (8.27) et (8.28) pour le cas $\beta = 2$.

Chapitre 9

Statistique d'extrêmes pour les matrices de Wishart

Après avoir présenté le modèle des matrices de Wishart dans la section 9.1, nous introduisons la famille de polynômes orthogonaux dans la section 9.2 pour étudier la statistique de la valeur propre minimale en suivant la méthode de Nadal et Majumdar [118]. À l'aide de cette analyse, nous retrouverons les comportements de la statistique de la valeur propre minimale [155] à N fini (section 9.3), dans la limite de *hard edge* (section 9.4) et dans la limite de *soft edge* (section 9.5). De plus, notre étude prouve une conjecture faite par Edelman *et al.* [47] (9.55).

9.1 Modèle des matrices de Wishart

Les matrices de Wishart (parfois appelées matrices de Laguerre ou encore matrices de Laguerre-Wishart) sont des matrices aléatoires possédant diverses applications. Les matrices de Wishart sont construites à partir de matrice dont les éléments sont gaussiens. Nous nous contenterons par la suite de ne parler que des matrices de Wishart complexes $\beta = 2$. Une telle matrice de Wishart X est une matrice complexe carrée $N \times N$ qui est formée par le produit $X = Y^\dagger Y$ avec Y une matrice rectangulaire $N \times M$ dont les éléments sont des gaussiennes complexes indépendantes $\mathcal{N}[0, 1] + i\mathcal{N}[0, 1]$ et $^\dagger Y$ la matrice hermitienne conjuguée de Y . On supposera par la suite $M \geq N$, quitte à inverser le rôle de M et N , et noterons $M = N + a$. Ces matrices ont été introduites par Wishart [165] dans le domaine des statistiques pour la méthode d'*analyse en composantes principales* [117, 146]. Désormais, les matrices de Wishart ont trouvé diverses applications : algorithmes de traitement d'images [73], génétique des populations [25], finance [15], météorologie et océanographie [134] ainsi qu'aux excursions répulsives que nous avons mentionnées dans la section 4.2.2 [140, 156]. Cet ensemble est également très étudié car, comme nous le verrons, il possède de multiples comportements très intéressants sur les bords du spectre. Cet ensemble de matrices aléatoires est, comme le sont les ensembles gaussiens (6.2), invariant par changement de base unitaire. On peut donc oublier la statistique des vecteurs propres et ne garder que celle des valeurs propres qui s'écrit dans ce cas [87]

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j}^N (\lambda_i - \lambda_j)^2 \exp \left(- \sum_{i=1}^N \lambda_i \right) \prod_{i=1}^N \lambda_i^a, \quad (9.1)$$

avec Z_N une constante de normalisation et où les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_N$ ne peuvent être que positives à cause de la définition $X = Y^\dagger Y$. A priori, $a = M - N$ est naturellement un entier positif. Il est néanmoins nécessaire de généraliser au cas où a est un réel pour

étudier certains problèmes, par exemple les excursions répulsives nécessitent d'avoir $a = 1/2$ [140, 156]. Les fluctuations des valeurs propres sont encore une fois caractérisées par les deux régimes que sont le régime de volume et le régime de bord. On peut réécrire cette loi jointe (9.1) comme un poids de Boltzmann

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) \propto \exp \left(- \sum_{i=1}^N \lambda_i + a \sum_{i=1}^N \log(\lambda_i) - \sum_{i \neq j} \ln(\lambda_i - \lambda_j) \right). \quad (9.2)$$

Une étude en ordre de grandeur de ce poids de Boltzmann, comme nous l'avons déjà fait pour le cas gaussien (6.9), permet de trouver que le comportement de volume des valeurs propres est d'ordre N . On peut étudier la densité de la valeur propre donnée par

$$\rho_0(\lambda, N) = \frac{1}{N} \sum_{i=1}^N \langle \delta(\lambda_i - \lambda) \rangle, \quad (9.3)$$

où la moyenne $\langle \dots \rangle$ correspond à la moyenne suivant la loi de probabilité jointe (9.1). Dans la limite thermodynamique de N grand, cette densité tend, dans le régime de volume $\lambda \sim \mathcal{O}(N)$, vers la loi de Marcenko-Pastur

$$\lim_{N \rightarrow \infty} N \rho_0(xN, N) = \rho_{\text{MP}}(x) = \frac{\sqrt{(x - x_-)(x_+ - x)}}{2\pi x}, \quad (9.4)$$

où les deux bornes du support sont données en fonction du paramètre $c = N/M \leq 1$ par

$$x_{\pm} = (c^{-1/2} \pm 1)^2. \quad (9.5)$$

Nous avons tracé dans la figure 9.1 la loi de Marcenko-Pastur (9.4). Dans le cas où $c < 1$, cas (a) sur la figure, la densité moyenne s'annule comme une racine au voisinage de chacun de ces bords (x_- et x_+). On s'attend donc à retrouver la classe d'universalité *soft edge*, à laquelle appartient le modèle des matrices gaussiennes, pour la statistique d'extrêmes (minimum et maximum) lorsque l'on considère des valeurs propres d'ordre $|\lambda_i - Nx_{\pm}| = \mathcal{O}(N^{1/3})$

$$\frac{1}{N} = \int_0^{\lambda_{\min}} d\lambda \rho_0(\lambda, N) \sim \frac{1}{N} \int_{Nx_-}^{\lambda_{\min}} d\lambda \sqrt{\frac{\lambda}{N} - x_-} \Rightarrow \lambda_{\min} - Nx_- \sim N^{1/3}. \quad (9.6)$$

On retrouve ainsi pour ce cas la loi de Tracy-Widom et le noyau d'Airy (en utilisant la bonne loi d'échelle) à la fois pour la statistique de bord proche du maximum et du minimum. Remarquons que pour obtenir ce cas, il est nécessaire de prendre la limite N vers l'infini et a vers l'infini avec le paramètre $c = 1/(1 + a/N)$ fini. Si le paramètre a reste fini, on a alors $c = 1$ qui correspond au cas (b) sur la figure 9.1. Dans ce cas, la loi de Marcenko-Pastur est

$$\rho_{\text{MP}}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}. \quad (9.7)$$

Proche du bord $x_+ = 4$, cette densité moyenne s'annule encore une fois comme une racine et l'on s'attend encore une fois à retrouver la classe d'universalité *soft edge* pour la statistique du maximum (Tracy-Widom et noyau d'Airy en utilisant la bonne loi d'échelle). Par contre, proche de $x_- = 0$, la densité moyenne possède une divergence (intégrable)

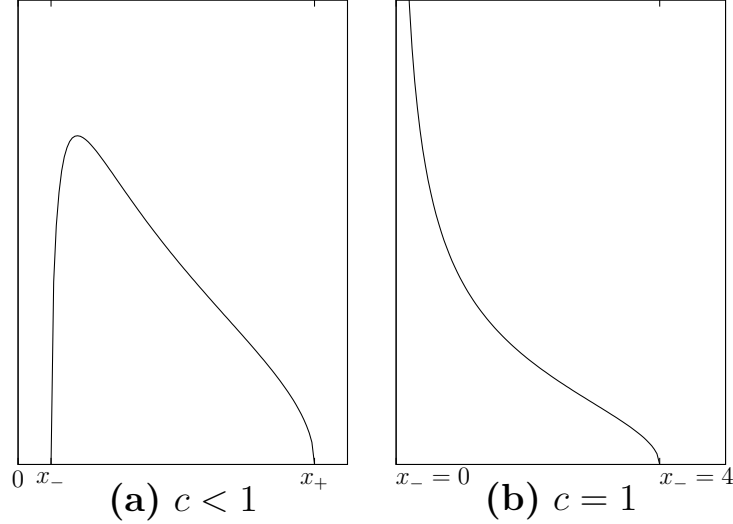


FIGURE 9.1: Loi de Marcenko-Pastur 9.4. **(a)** Dans le cas où $c < 1$, la loi de Marcenko-Pastur s'annule comme une racine au voisinage de chacun de ces bords. On s'attend donc à retrouver le classe d'universalité de *soft edge* pour les deux régimes de bords. **(b)** Dans le cas où $c < 1$, la loi de Marcenko-Pastur s'annule comme une racine au voisinage de $x_+ = 4$ mais possède une divergence intégrable en $x_- = 0$. On s'attend donc à retrouver le classe d'universalité de *soft edge* dans le régime de bord vers $x_+ = 4$ et à trouver une nouvelle classe d'universalité, la classe d'universalité de *hard edge* dans le régime de bord vers $x_- = 0$.

$\rho_{\text{MP}}(x) \sim x^{-1/2}$. On s'attend alors à observer une nouvelle classe d'universalité au voisinage de la valeur propre minimale λ_{\min} , on parle d'universalité *hard edge*. La loi d'échelle de ce régime de bord est alors donnée par

$$\frac{1}{N} = \int_0^{\lambda_{\min}} d\lambda \rho_0(\lambda, N) \sim \frac{1}{N} \int_0^{\lambda_{\min}} d\lambda \sqrt{\frac{N}{\lambda}} \Rightarrow \lambda_{\min} \sim N^{-1}. \quad (9.8)$$

Nous verrons que la statistique de la valeur propre minimale $\lambda_{\min} \sim N^{-1}$ est donnée par un système faisant intervenir une équation de Painlevé III. Nous cherchons alors par la suite à étudier la distribution de la valeur propre minimale

$$F_N(t) = \Pr. [\lambda_{\min} = \min_{1 \leq i \leq N} \lambda_i \geq t] \quad (9.9)$$

$$= \int_t^\infty d\lambda_1 \int_t^\infty d\lambda_2 \cdots \int_t^\infty d\lambda_N P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N), \quad (9.10)$$

dans la limite thermodynamique de grand N pour les deux classes d'universalité qui sont de *hard edge* et de *soft edge*.

9.2 Polynômes orthogonaux

Pour étudier la statistique de la valeur propre minimale, on introduit encore une fois [118] la famille étagée de polynômes orthogonaux unitaires

$$\begin{cases} \langle \pi_k | \pi_{k'} \rangle = \int_t^\infty e^{-\lambda} \lambda^a \pi_k(\lambda) \pi_{k'}(\lambda) d\lambda = h_k \delta_{k,k'}, \\ \pi_k(\lambda) = \lambda^k + \zeta_k \lambda^{k-1} + \dots \end{cases} \quad (9.11)$$

qui sont des polynômes de la variable λ qui dépendent des paramètres t et a . Nous avons introduit pour ce cas les variables ζ_k qui sont les seconds coefficients des polynômes. Quand le paramètre t tend vers 0, ces polynômes orthogonaux dégénèrent en polynômes de Laguerre généralisés (22). En utilisant ces polynômes orthogonaux, il est standard de réécrire, comme nous l'avons fait dans le cas gaussien (7.23), la distribution de la valeur propre minimale [65, 113, 118]

$$F_N(t) = \frac{N!}{Z_N} \prod_{k=0}^{N-1} h_k. \quad (9.12)$$

Comme pour le cas gaussien (7.11), on peut écrire une relation de récurrence à 3 termes pour ces polynômes [147]

$$\lambda \pi_k = \pi_{k+1} + S_k \pi_k + R_k \pi_{k-1}. \quad (9.13)$$

Une étude approfondie [128] des variables S_k , R_k , h_k et ζ_k permet de montrer que les variables S_k et R_k vérifient les relations de récurrence dite de Schlesinger [102]

$$\begin{cases} S_k - R_{k+1} + R_k &= t \partial_t S_k, \\ 2 - S_{k+1} + S_k &= t \frac{\partial_t R_{k+1}}{R_{k+1}}, \end{cases} \quad (9.14)$$

avec les conditions initiales

$$\begin{cases} S_0 &= -t \partial_t \log \left(\int_t^\infty e^{-\lambda} \lambda^a \right) + a + 1, \\ R_0 &= 0. \end{cases} \quad (9.15)$$

Tandis que les variables h_k et ζ_k sont liées aux quantités R_k et S_k par les relations

$$R_k = \frac{h_k}{h_{k-1}} \quad (9.16)$$

$$S_k = -t \partial_t \log h_k + 2k + a + 1. \quad (9.17)$$

$$\zeta_k = \zeta_{k-1} - S_{k-1} = - \sum_{i=0}^{k-1} S_i. \quad (9.18)$$

Il n'est pas difficile de montrer que ce système d'équations détermine de façon unique les différentes variables S_k , R_k , h_k et ζ_k . En analysant ce système d'équations [8, 128], il est possible de montrer que ces variables satisfont également les équations de récurrence de Laguerre-Freud

$$\begin{cases} R_{k+2} - R_k &= S_{k+1}(2k + 4 + a + t - S_{k+1}) - S_k(2k + a + t - S_k) - 2t, \\ S_{k+1}(S_{k+1} - t) &= R_{k+1}(2k + 1 + a + t - S_{k+1} - S_k) \\ &\quad - R_{k+2}(2k + 5 + a + t - S_{k+2} - S_{k+1}). \end{cases} \quad (9.19)$$

Il est possible, à partir de ces différents résultats de réécrire la distribution de la valeur propre minimale (9.12) sous la forme

$$t \partial_t \log (F_N(t)) = \zeta_N + N(N + a). \quad (9.20)$$

On introduit alors les grandeurs

$$\begin{cases} \theta_k &= 2k + 1 + a - S_k, \\ \omega_k &= -R_k - \zeta_k. \end{cases} \quad (9.21)$$

On peut montrer [68, 128] que les équations de Laguerre-Freud (9.19) se réécrivent

$$\omega_k^2 = R_k \theta_k \theta_{k-1} \quad (9.22)$$

$$(\omega_k - kt)(\omega_k - (k + a)t) = R_k(\theta_k + t)(\theta_{k-1} + t). \quad (9.23)$$

On trouve alors l'équation

$$\omega_k^2 - \theta_k(\theta_k - a - 2k + t)\omega_k - \theta_k(kt(k + a) + (\theta_k + t)\zeta_k) = 0, \quad (9.24)$$

qui, comme nous allons le voir, permet facilement d'étudier la limite thermodynamique de grand N dans le régime de *hard edge*.

9.3 Équation de Painlevé V à N fini

Nous allons dans cette partie dériver l'équation de Painlevé V qui gouverne la statistique de la valeur propre minimale lorsque la taille N de la matrice est finie. Pour cela, nous posons la fonction

$$H_N = t\partial_t \log(F_N(t)) = N(N + a) + \zeta_N. \quad (9.25)$$

En sommant la première équation de Schlesinger (9.14), on trouve

$$\sum_{k=0}^{N-1} S_k + \sum_{k=0}^{N-1} (R_k - R_{k-1}) = t\partial_t \sum_{k=0}^{N-1} S_k. \quad (9.26)$$

En utilisant les relations (9.18) et (9.21) couplées aux conditions initiales (9.15), on trouve alors l'équation

$$-t\partial_t \zeta_k = -R_k - \zeta_k = \omega_k. \quad (9.27)$$

Cette équation permet, en dérivant (9.25) par rapport à t , d'écrire

$$t\partial_t H_N = -\omega_N = \zeta_N + R_N = H_N - N(N + a) + R_N. \quad (9.28)$$

En redérivant cette équation par rapport à la variable t , on trouve

$$\partial_t R_N = t\partial_t^2 H_N. \quad (9.29)$$

En utilisant la seconde équation de Schlesinger (9.14), on trouve avec la définition (9.21)

$$t\partial_t R_k = R_k(2 - S_k - S_{k-1}) \quad (9.30)$$

$$= R_k(\theta_k - \theta_{k-1}). \quad (9.31)$$

À partir du résultat (9.22), il est possible de réécrire cette dernière équation à l'aide de quantités prises au même indice

$$t\partial_t R_k = R_k \theta_k - \frac{\omega_k^2}{\theta_k}. \quad (9.32)$$

En combinant les équations (9.29) et (9.32), on trouve alors

$$t^2\partial_t^2 H_N = R_N \theta_N - \frac{\omega_N^2}{\theta_N}. \quad (9.33)$$

De plus, en utilisant l'équation (9.24) avec la définition de ω_k (9.21) pour éliminer la grandeur ζ_k , on trouve

$$k(k+a)t - (2k+a)\omega_k - tR_k = R_k\theta_k + \frac{\omega_k^2}{\theta_k}. \quad (9.34)$$

En sommant et soustrayant ces deux équations (9.33) et (9.34), on trouve alors (au rang $k = N$)

$$2\frac{\omega_N^2}{\theta_N} = N(N+a)t - (2N+a)\omega_N - tR_N - t^2\partial_t^2 H_N \quad (9.35)$$

$$2R_N\theta_N = N(N+a)t - (2N+a)\omega_N - tR_N + t^2\partial_t^2 H_N. \quad (9.36)$$

En multipliant terme à terme ces deux équations, on obtient

$$4R_N\omega_N^2 = (N(N+a)t - (2N+a)\omega_N - tR_N)^2 - (t^2\partial_t^2 H_N)^2. \quad (9.37)$$

Il ne reste qu'à éliminer dans cette équation les grandeurs ω_N et R_N en utilisant (9.28) pour trouver l'équation de Painlevé V

Équation de Painlevé V

$$(t\partial_t^2 H_N)^2 = 4(\partial_t H_N)^2(H_N - N(N+a) - t\partial_t H_N) + ((2N+a-t)\partial_t H_N + H_N)^2. \quad (9.38)$$

9.4 Étude du *hard edge*

On cherche désormais à étudier la limite asymptotique de *hard edge*, c'est-à-dire la limite thermodynamique N grand avec un paramètre a fini (et donc c tendant vers 1). Nous avons vu (9.8) que dans ce régime, la valeur propre minimale était d'ordre N^{-1} . Ainsi, pour obtenir le régime de bord, il est nécessaire d'analyser la limite d'échelle

$$N \rightarrow \infty, \quad t \rightarrow 0, \quad x = Nt \text{ fini}. \quad (9.39)$$

Une méthode pour analyser ce comportement est d'utiliser ce régime d'échelle dans l'équation de Painlevé V (9.38) [68, 155]. Cette méthode permet bien de trouver l'équation de Painlevé III propre au régime de *hard edge* mais il est difficile d'étudier les premières corrections. L'analyse de l'équation (9.24) dans le bon régime d'échelle, permet quant à elle de trouver les premières corrections et de vérifier une conjecture de Edelman *et al.* [47]. Pour étudier la limite d'échelle (9.39), nous commençons par étudier le cas $t = 0$. Dans ce cas, la famille de polynômes orthogonaux est la famille des polynômes de Laguerre généralisés (22) et il est donc facile de calculer les différentes grandeurs

$$\begin{cases} \pi_k(\lambda)|_{t=0} &= L_k^{(a)}(\lambda)k!(-1)^k, \\ h_k|_{t=0} &= (k+a)!k!, \\ R_k|_{t=0} &= k(k+a), \\ S_k|_{t=0} &= 2k+a+1, \\ \zeta_k|_{t=0} &= -k(k+a). \end{cases} \quad (9.40)$$

En dérivant la définition des polynômes orthogonaux (9.11) par rapport à la variable t , on trouve, car les polynômes sont unitaires

$$\partial_t h_k = -e^{-\alpha t} t^a \pi_k^2(t). \quad (9.41)$$

On peut alors réintégrer cette équation lorsque t est proche de 0 pour trouver à l'aide des résultats (9.40)

$$h_k = (k+a)!k! - \frac{t^{a+1}(k+a)!^2}{(a+1)!a!} + o(t^{a+1}). \quad (9.42)$$

En utilisant le comportement d'échelle (9.39), on peut réécrire cette équation comme

$$h_N = (N+a)!N! \left[1 - \frac{1}{N} \left(\frac{x^{a+1}}{a!(a+1)!} + o(x^{a+1}) \right) + o\left(\frac{1}{N}\right) \right], \quad (9.43)$$

ce qui suggère l'ansatz

$$h_N = (N+a)!N! \left[1 + \frac{1}{N} f(x) + o\left(\frac{1}{N}\right) \right], \quad (9.44)$$

où l'on a introduit une fonction $f(x)$ dont on connaît pour l'instant uniquement le comportement à petit x

$$f(x) = -\frac{x^{a+1}}{a!(a+1)!} + o(x^{a+1}). \quad (9.45)$$

En utilisant cet ansatz dans les relations (9.16) et (9.17) on trouve

Comportement asymptotique de R_n et S_n dans la limite de *hard edge*

$$\begin{cases} R_N &= N(N+a) + xf'(x) - f(x) + o(1), \\ S_N &= 2N + a + 1 - \frac{1}{N}xf'(x) + o\left(\frac{1}{N}\right). \end{cases} \quad (9.46)$$

Mentionnons que le développement de h_{N-1} fait apparaître dans la fonction f l'argument $(N-1)t = x(1-1/N)$. Ces différents résultats suggèrent alors les développements

$$\begin{cases} R_N &= N(N+a) + \sum_{i=0}^j r_i(x)N^{-i} + o(N^{-j}), \\ S_N &= 2N + a + 1 + \sum_{i=1}^j s_i(x)N^{-i} + o(N^{-j}). \end{cases} \quad (9.47)$$

où les premiers termes sont donnés par l'équation (9.46). En injectant ces développements (9.47) dans les équations de Schlesinger (9.14), on peut calculer terme à terme les différents ordres. Par la suite, nous aurons besoin des deux premiers termes

$$\begin{cases} R_N &= N(N+a) + xf'(x) - f(x) + \frac{a}{2N}x^2f''(x) + o\left(\frac{1}{N}\right) \\ S_N &= 2N + a + 1 - \frac{1}{N}xf'(x) - \frac{a+1}{2N^2}x^2f''(x) + o\left(\frac{1}{N^2}\right). \end{cases} \quad (9.48)$$

En utilisant l'équation (9.18), on peut calculer ζ_N avec la même précision

$$\zeta_N = -(N+a)N + f(x) + \frac{a}{2N}xf'(x) + o\left(\frac{1}{N}\right). \quad (9.49)$$

À partir de la définition (9.21), on peut obtenir les développements de θ_N et ω_N que l'on injecte alors dans l'équation (9.24) (avec $t = x/N$), pour obtenir l'équation de Painlevé III

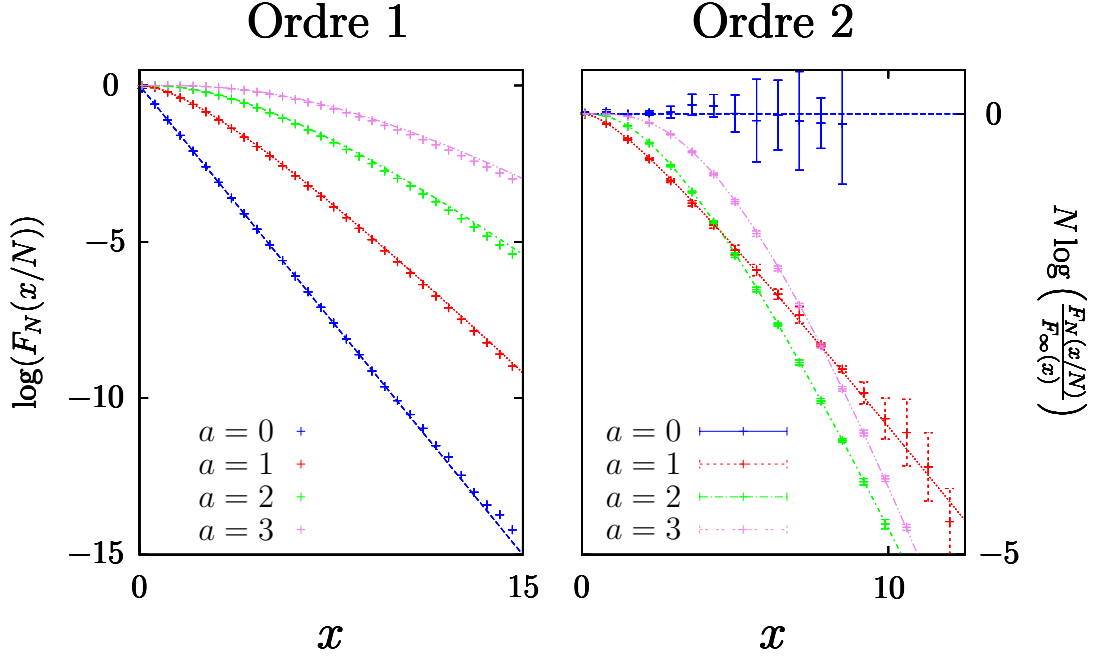


FIGURE 9.2: Comparaison entre la formule (9.53) (lignes) et des données obtenues en générant 3.10^7 matrices de Wishart de taille $N = 50$ (points) dans le cas de différents a entiers. Le graphique de gauche compare $\log(F_N(x/N))$ à la formule de $\log(F_\infty(x))$ donnée en équation (9.54). Le graphique de droite analyse les corrections à cette formule en comparant $N \log \left(\frac{F_N(x/N)}{F_\infty(x)} \right)$ à $\frac{a}{2} f(x)$ (9.53).

Équation de Painlevé III

$$(xf'')^2 + 4f'(1+f')(xf' - f) = (af')^2. \quad (9.50)$$

On retrouve, d'une autre façon, un résultat obtenu par Tracy et Widom [155]. Notons que dans leur papier, la fonction $\sigma(s)$ est donnée par $-f(s/4)$. Dans le cas où a est un entier, il est possible de résoudre cette équation avec la condition aux limites (9.45) [66] et d'exprimer la fonction f à partir de la fonction de Bessel modifiée de première espèce (16)

$$f(x) = -x \frac{\det [I_{j-k+2}(2\sqrt{x})]_{1 \leq j, k \leq a}}{\det [I_{j-k}(2\sqrt{x})]_{1 \leq j, k \leq a}}. \quad (9.51)$$

En utilisant le lien entre la distribution de la valeur propre minimale et le coefficient ζ_n (9.20), on trouve

$$x \partial_x \log \left(F_N \left(\frac{x}{N} \right) \right) = f(x) + \frac{a}{2N} x f'(x) + o \left(\frac{1}{N} \right). \quad (9.52)$$

En utilisant la condition initiale $F_N(0) = 1$, on peut intégrer cette équation

$$F_N \left(\frac{x}{N} \right) = \exp \left(\int_0^x \frac{f(u)}{u} du + \frac{a}{2N} f(x) + o \left(\frac{1}{N} \right) \right). \quad (9.53)$$

Ainsi, au premier ordre, la distribution de la valeur propre minimale est donc donnée par

Distribution de la valeur propre minimale

$$\lim_{N \rightarrow \infty} F_N \left(\frac{x}{N} \right) = F_\infty(x), \quad F_\infty(x) = \exp \left(\int_0^x \frac{f(u)}{u} du \right), \quad (9.54)$$

où la fonction f vérifie l'équation de Painlevé III (9.50) [155]. Comme le montre la partie gauche de la figure 9.2, cette formule colle bien aux données obtenues en simulant un grand nombre de matrices de Wishart pour les cas a entier en utilisant le résultat (9.51). Au second ordre, le résultat (9.53), permet de prouver une conjecture faite par Edelman *et al.* [47]

Correction à la distribution de la valeur propre minimale

$$\lim_{N \rightarrow \infty} N \left(F_N \left(\frac{x}{N} \right) - F_\infty(x) \right) = \frac{a}{2} x F'_\infty(x). \quad (9.55)$$

La partie droite de la figure 9.2 étudie le second ordre du développement de F_N en traçant la grandeur $N \log(F_N(x/N)/F_\infty(x))$ qui d'après (9.53) est égale à $\frac{a}{2} f(x)$ que l'on peut calculer en utilisant (9.51).

9.5 Étude du *soft edge*

On cherche dans cette partie à étudier la limite de *soft edge*. Nous avons vu dans la section 9.1 que cette limite s'obtient en prenant a d'ordre N pour que le paramètre $c = 1/(1 + a/N)$ tendent vers une valeur inférieure à 1 dans la limite de N grand. On pose donc pour la suite $a = \alpha N$. Dans cette limite, la moyenne de la valeur propre minimale est donné par le bord du spectre, au premier ordre $\langle \lambda_{\min} \rangle \sim Nx_-$ (9.5) avec des fluctuations autour de cette moyenne d'ordre $N^{1/3}$ (9.6). On pose donc

$$t = Nx_- - \frac{N^{1/3}}{m} x, \quad (9.56)$$

où m est un paramètre qui ne dépend que de α . En utilisant ce régime d'échelle, la distribution de la valeur propre minimale peut être développée sous la forme

$$F_N(t) = f_0 \left(m \frac{Nx_- - t}{N^{1/3}} \right) + O(N^{-1/3}). \quad (9.57)$$

En utilisant ce développement on trouve que la quantité H_N admet le développement

$$\begin{aligned} H_N = t \partial_t \log F_N(t) &= -mx_- \frac{\partial_x f_0(x)}{f_0(x)} N^{2/3} + \dots \\ &= -mx_- g_0(x) N^{2/3} + \dots \end{aligned} \quad (9.58)$$

Il ne reste alors qu'à injecter ce comportement dans l'équation de Painlevé V (9.38), en posant $m = \frac{(1+\alpha)^{1/6}}{(\sqrt{1+\alpha}-1)^{4/3}}$ pour simplifier le résultat, pour trouver

$$\left(\partial_x^2 g_0(x) \right)^2 + 4 \partial_x g_0(x) \left[(\partial_x g_0(x))^2 - x \partial_x g_0(x) + g_0(x) \right] = 0. \quad (9.59)$$

La solution de cette équation est alors $g_0(x) = \int_x^\infty q^2(u) du$ avec q la solution de Hastings-McLeod, de l'équation de Painlevé II (6.20). Finalement, on retrouve bien la distribution de Tracy-Widom (6.21)

— Distribution de la valeur propre minimale dans la limite de *soft edge* —

$$\lim_{N \rightarrow \infty} F_N \left(Nx_- - N^{1/3} \frac{(\sqrt{1+\alpha} - 1)^{4/3}}{(1+\alpha)^{1/6}} x \right) = \exp \left(- \int_x^\infty du (u-x) q^2(u) \right) = \mathcal{F}_2(x). \quad (9.60)$$

Appendices

Annexe F

Analyse du régime de bord pour des matrices gaussiennes unitaires

Nous avons vu dans le texte que la fonction $\tilde{f}(\tilde{r}, x)$ vérifient une équation de Schrödinger supersymétrique (7.41). En effet, en utilisant l'équation de Painlevé II (6.20) que vérifie la solution de Hasting-McLeod $q(x)$, on peut réécrire

$$x + 2q^2(x) = \frac{q''(x)}{q(x)} = \frac{d}{dx} \left(\frac{q'(x)}{q(x)} \right) + \left(\frac{q'(x)}{q(x)} \right)^2 = -Q'(x) + Q^2(x) , \quad (\text{F.1})$$

$$\text{avec } Q(x) = -\frac{q'(x)}{q(x)} . \quad (\text{F.2})$$

On peut alors introduire [63, 126] la fonction $\tilde{g}(\tilde{r}, x)$ telle que

$$\partial_x \tilde{g}(\tilde{r}, x) = r \tilde{f}(\tilde{r}, x) + Q(x) \tilde{g}(\tilde{r}, x) \quad (\text{F.3})$$

$$\partial_x \tilde{f}(\tilde{r}, x) = -Q(x) \tilde{f}(\tilde{r}, x) - \tilde{g}(\tilde{r}, x) . \quad (\text{F.4})$$

En effet, en dérivant la seconde équation, on retrouve bien l'équation de Schrödinger (7.41) en utilisant la forme supersymétrique (F.1).

$$\begin{aligned} \partial_x^2 \tilde{f}(\tilde{r}, x) &= -Q'(x) \tilde{f}(\tilde{r}, x) - Q(x) \partial_x \tilde{f}(\tilde{r}, x) - \partial_x \tilde{g}(\tilde{r}, x) \\ &= -Q'(x) \tilde{f}(\tilde{r}, x) - Q(x) (-Q(x) \tilde{f}(\tilde{r}, x) - \tilde{g}(\tilde{r}, x)) - (r \tilde{f}(\tilde{r}, x) + Q(x) \tilde{g}(\tilde{r}, x)) \\ &= (-Q'(x) + Q^2(x)) \tilde{f}(\tilde{r}, x) - r \tilde{f}(\tilde{r}, x) . \end{aligned} \quad (\text{F.5})$$

On peut réécrire ces équations sous la forme matricielle

$$\frac{\partial}{\partial x} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{B}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \frac{q'(x)}{q(x)} & -1 \\ \tilde{r} & -\frac{q'(x)}{q(x)} \end{pmatrix} . \quad (\text{F.6})$$

Remarquons que la première équation (F.3) peut s'intégrer, en utilisant le comportement asymptotique (7.41), sous la forme [126]

$$q(x) \tilde{g}(\tilde{r}, x) = -r \int_x^\infty q(u) \tilde{f}(\tilde{r}, u) du . \quad (\text{F.7})$$

On peut alors montrer [63, 126] que la paire de fonction $(\tilde{f}(\tilde{r}, x), \tilde{g}(\tilde{r}, x))$ forment une paire de Lax qui vérifie

$$\frac{\partial}{\partial \tilde{r}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{A}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad \tilde{\mathbf{A}} = \begin{pmatrix} -\frac{q'(x)}{q(x)} & 1 + q^2(x)/\tilde{r} \\ -\tilde{r} - \frac{\int_x^\infty q^2(u) du}{q^2(x)} & \frac{q'(x)}{q(x)} \end{pmatrix} . \quad (\text{F.8})$$

En utilisant que $\tilde{f}(\tilde{r}, x) = 2^{-1/6}G(\tilde{r}, x)$ dans la forme limite du noyau (7.42) avec l'équation (F.4), on trouve que celui-ci se réécrit

$$K_N \left(y - \frac{\tilde{r}}{\sqrt{2}}N^{-1/6}, y - \frac{\tilde{r}'}{\sqrt{2}}N^{-1/6} \right) \underset{N \rightarrow \infty}{\sim} N^{1/6} 2^{5/6} \frac{\tilde{f}(\tilde{r}, x)\tilde{g}(\tilde{r}', x) - \tilde{f}(\tilde{r}', x)\tilde{g}(\tilde{r}, x)}{\pi(\tilde{r} - \tilde{r}')} . \quad (\text{F.9})$$

Pour calculer la densité à partir du maximum $\rho_{\max}(r, N)$ (7.26), il est nécessaire de connaître les comportements de $K_N(y - r, y - r)$ et $K_N(y, y - r)$. En prenant $\tilde{r} = 0$ dans l'équation de Schrödinger (7.41), on trouve que la solution possédant le bon comportement à grand x est

$$\tilde{f}(\tilde{r} = 0, x) = 2^{-1/6}\sqrt{\pi}q(x) . \quad (\text{F.10})$$

En utilisant alors (F.7), on trouve que $\tilde{g}(0, x) = 0$ et donc avec l'expression (F.9)

$$K_N \left(y - \frac{\tilde{r}}{\sqrt{2}}N^{-1/6}, y \right) \underset{N \rightarrow \infty}{\sim} -N^{1/6} 2^{2/3} \frac{\tilde{g}(\tilde{r}, x)q(x)}{\sqrt{\pi}\tilde{r}} . \quad (\text{F.11})$$

En utilisant la règle de L'Hospital, on peut calculer le noyau au point coïncidant en utilisant les équations (F.8)

$$K_N(y - r, y - r) \underset{N \rightarrow \infty}{\sim} \frac{2^{5/6}N^{1/6}}{\pi} \left[\partial_{\tilde{r}}\tilde{f}(\tilde{r}, x)\tilde{g}(\tilde{r}, x) - \tilde{f}(\tilde{r}, x)\partial_{\tilde{r}}\tilde{g}(\tilde{r}, x) \right] \quad (\text{F.12})$$

$$= \frac{2^{5/6}N^{1/6}}{\pi} \left[\left(\tilde{r} + \frac{\int_x^\infty q^2(u)du}{q^2(x)} \right) \tilde{f}^2 - 2\frac{q'(x)}{q(x)}\tilde{f}\tilde{g} + \left(1 + \frac{q^2(x)}{\tilde{r}} \right) \tilde{g}^2 \right] , \quad (\text{F.13})$$

où l'on a noté $\tilde{f} = \tilde{f}(\tilde{r}, x)$ et $\tilde{g} = \tilde{g}(\tilde{r}, x)$. En utilisant les résultats à $\tilde{r} = 0$, on trouve alors

$$K_N(y, y) \underset{N \rightarrow \infty}{\sim} \sqrt{2}N^{1/6} \int_x^\infty q^2(u)du . \quad (\text{F.14})$$

On retrouve donc en utilisant l'identité (7.25) que la distribution de la valeur propre maximale est donnée par la loi de Tracy-Widom (6.21) [154]

$$\log F_N(y) \underset{N \rightarrow \infty}{\rightarrow} \log \mathcal{F}_2(\sqrt{2}N^{1/6}(y - \sqrt{2N})) . \quad (\text{F.15})$$

En utilisant les expressions (F.11), (F.12) et (F.15) dans la densité à partir du maximum (7.26), on trouve finalement le comportement de bord

$$\begin{aligned} \rho_{\max}(r, N) &= N^{-5/6}\sqrt{2}\tilde{\rho}_{\text{edge}}(\sqrt{2}N^{1/6}r) , \\ \tilde{\rho}_{\text{edge}}(\tilde{r}) &= \frac{2^{1/3}}{\pi} \int_{-\infty}^\infty \left[R \left(\left(\tilde{r} + \frac{R}{q^2} \right) \tilde{f}^2 - 2\frac{q'}{q}\tilde{f}\tilde{g} + \left(1 + \frac{q^2}{\tilde{r}} \right) \tilde{g}^2 \right) - \frac{1}{\tilde{r}^2}q^2\tilde{g}^2 \right] \mathcal{F}_2 dx , \end{aligned} \quad (\text{F.16})$$

où l'on a utilisé les notations abrégées $q = q(x)$, $R = R(x)$, $\mathcal{F}_2 = \mathcal{F}_2(x)$, $\tilde{f} = \tilde{f}(\tilde{r}, x)$ et $\tilde{g} = \tilde{g}(\tilde{r}, x)$ avec

$$R(x) = \int_x^\infty q^2(u)du = \frac{\mathcal{F}_2'(x)}{\mathcal{F}_2(x)} . \quad (\text{F.17})$$

Cette expression (F.16) peut être simplifiée en utilisant l'identité obtenue à partir des équations (F.8)

$$\partial_x \left(\left(\tilde{r} + \frac{R}{q^2} \right) \tilde{f}^2 - 2\frac{q'}{q}\tilde{f}\tilde{g} + \left(1 + \frac{q^2}{\tilde{r}} \right) \tilde{g}^2 \right) = -\tilde{f}^2 . \quad (\text{F.18})$$

On peut donc réécrire la densité à partir du maximum dans le régime de bord comme

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[R(x) \int_x^{\infty} \tilde{f}^2(\tilde{r}, u) \, du - q^2(x) \frac{\tilde{g}(\tilde{r}, x)}{\tilde{r}^2} \right] \mathcal{F}_2(x) \, dx . \quad (\text{F.19})$$

Cette expression peut encore être simplifiée pour trouver le résultat (7.43) en utilisant une intégration par parties à l'aide des identités (F.7) et (F.17)

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(\tilde{r}, x) - \frac{q^2(x)}{\tilde{r}^2} \tilde{g}^2(\tilde{r}, x) \right] \mathcal{F}_2(x) \, dx \quad (\text{F.20})$$

$$= \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(\tilde{r}, u) \, du \right)^2 \right] \mathcal{F}_2(x) \, dx . \quad (\text{F.21})$$

Annexe G

Densité à partir du maximum et premier gap pour des matrices de Wishart $a = 0$

À partir des développements (9.46), il est sûrement possible de reproduire l'étude que l'on a faite dans le chapitre 7 pour les matrices gaussiennes unitaires. Nous nous concentrerons uniquement ici sur le cas des matrices de Wishart $a = 0$. Dans ce cas, nous verrons qu'il est facile de calculer la densité à partir du minimum ainsi que le premier gap.

G.1 Modèle de Wishart pour $a = 0$

Dans le cas où $a = 0$, la loi jointe des valeurs propres devient (9.1)

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j} (\lambda_i - \lambda_j)^2 \exp \left(- \sum_{i=1}^N \lambda_i \right). \quad (\text{G.1})$$

En particulier, cette loi vérifie

$$P_{\text{joint}}(\lambda_1 + t, \lambda_2 + t, \dots, \lambda_N + t) = e^{-Nt} P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N). \quad (\text{G.2})$$

En utilisant cette relation (G.2), il est facile de calculer la distribution de la valeur propre minimale (9.9) [64]

$$F_N(t) = \int_t^\infty d\lambda_1 \int_t^\infty d\lambda_2 \cdots \int_t^\infty d\lambda_N P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) \quad (\text{G.3})$$

$$= \int_0^\infty dz_1 \int_0^\infty dz_2 \cdots \int_0^\infty dz_N P_{\text{joint}}(z_1, z_2, \dots, z_N) e^{-Nt} \quad (\text{G.4})$$

$$= e^{-Nt}, \quad (\text{G.5})$$

où la dernière ligne est obtenue en utilisant la normalisation de la loi jointe. Ainsi, dans le cas $a = 0$, la valeur propre minimale est d'ordre $O(N^{-1})$ et est distribuée suivant une loi exponentielle. On retrouve évidemment les fluctuations du *hard edge* (9.8). Notons que la distribution exponentielle est bien en accord avec le résultat (9.54) avec la fonction $f(x) = -x$ (9.51), les corrections (9.55) étant pour ce cas nulles. Via l'identité (G.2), il est intéressant d'utiliser dans ce cas la famille des polynômes orthogonaux de Laguerre définie par

$$\int_0^\infty dx L_k(x) L_{k'}(x) e^{-x} = \delta_{k,k'}. \quad (\text{G.6})$$

Le noyau est alors donné par, en utilisant la formule de Christoffel-Darboux

$$K_N(x, y) = \sum_{k=0}^{N-1} L_k(x) L_k(y) e^{-\frac{x+y}{2}} \quad (\text{G.7})$$

$$= N \frac{L_{N-1}(x) L_N(y) - L_{N-1}(y) L_N(x)}{x - y} e^{-\frac{x+y}{2}}. \quad (\text{G.8})$$

En particulier, le noyau pris au point coïncident est, en utilisant les propriétés des polynômes de Laguerre

$$\begin{aligned} K_N(x, x) &= N \left[L'_{N-1}(x) L_N(x) - L_{N-1}(x) L'_N(x) \right] e^{-x} \\ &= N \left[(N-1) (L_{N-1}(x) - L_{N-2}(x)) L_N(x) - L_{N-1}(x) N (L_N(x) - L_{N-1}(x)) \right] \frac{e^{-x}}{x}. \end{aligned} \quad (\text{G.9})$$

En utilisant la formule de Plancherel-Rotach, on trouve

$$e^{-\frac{x}{2}} L_N(x) \underset{N \rightarrow \infty}{=} J_0 \left(2 \sqrt{\left(N + \frac{1}{2} \right) x} \right) + o(N^{-1}). \quad (\text{G.10})$$

On trouve alors que le noyau (G.8), dans le régime de bord, est donné par le noyau de Bessel [145, 155]

$$\frac{1}{N} K_N \left(\frac{x}{N}, \frac{y}{N} \right) \underset{N \rightarrow \infty}{\rightarrow} f_K(x, y) = \frac{\sqrt{x} J_0(2\sqrt{y}) J_1(2\sqrt{x}) - \sqrt{y} J_0(2\sqrt{x}) J_1(2\sqrt{y})}{x - y}, \quad (\text{G.11})$$

où les fonctions J_0 et J_1 sont les deux premières fonctions de Bessel de première espèce (15). Notons qu'il est similaire d'utiliser la famille des polynômes introduite au chapitre (9.11) qui sont donnés par des polynômes de Laguerre translatés.

G.2 Densité de valeurs propres

En utilisant les résultats classiques de la méthode des polynômes orthogonaux, on peut calculer la densité de valeur propre à partir du noyau [113]

$$\rho_0(x, N) = \frac{1}{N} \left\langle \sum_{i=0}^N \delta(x - \lambda_i) \right\rangle = \frac{1}{N} K_N(x, x). \quad (\text{G.12})$$

Ainsi, à partir de la formule (G.9), la densité de valeur propre est

$$\rho_0(x, N) = \frac{e^{-x}}{x} \left(N L_{N-1}(x)^2 - L_{N-1}(x) L_N(x) - (N-1) L_{N-2}(x) L_N(x) \right). \quad (\text{G.13})$$

Dans la limite de N grand, on retrouve bien le régime de volume de Marcenko-Pastur (9.7) avec $c = 1$

$$\lim_{N \rightarrow \infty} N \rho_0(Nx, N) = \rho_{\text{MP}}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}. \quad (\text{G.14})$$

Il est également possible d'étudier le régime de bord à partir de (G.11)

$$\lim_{N \rightarrow \infty} \rho_0 \left(\frac{r}{N}, N \right) = \rho_{0,E}(r) = f_K(r, r) = J_0(2\sqrt{r})^2 + J_1(2\sqrt{r})^2. \quad (\text{G.15})$$

En particulier, les limites asymptotiques sont données par

$$\rho_{0,E}(r) \sim \begin{cases} 1 - r, & r \rightarrow 0 \\ \frac{1}{\pi\sqrt{r}}, & r \rightarrow \infty, \end{cases} \quad (\text{G.16})$$

le comportement à r grand se raccordant parfaitement avec la loi de Marcenko-Pastur $c = 1$ (9.7).

G.3 Densité à partir de la valeur propre minimale

Nous étudions alors la densité à partir de la valeur propre minimale définie, de manière analogue à (7.1), par

$$\rho_{\max}(x, N) = \frac{1}{N-1} \left\langle \sum_{\substack{i=0 \dots N \\ i \neq i_{\min}}} \delta(\lambda_{\min} + x - \lambda_i) \right\rangle, \quad (\text{G.17})$$

où l'on a noté i_{\min} l'indice de la valeur propre minimale $\lambda_{i_{\min}} = \lambda_{\min}$. En utilisant l'identité (G.2), il est possible de donner une expression de cette densité à partir de la valeur propre minimale en termes du noyau (G.8) [113]

$$\begin{aligned} \rho_{\max}(x, N) &= N \int_0^\infty d\lambda_{\min} \int_{\lambda_{\min}}^\infty d\lambda_3 \cdots \int_{\lambda_{\min}}^\infty d\lambda_N P_{\text{joint}}(\lambda_{\min}, \lambda_{\min} + x, \lambda_3, \dots, \lambda_N) \\ &= N \int_0^\infty d\lambda_{\min} e^{-N\lambda_{\min}} \int_0^\infty dz_3 \cdots \int_0^\infty dz_N P_{\text{joint}}(0, x, z_3, \dots, z_N) \\ &= \int_0^\infty dz_3 \cdots \int_0^\infty dz_N P_{\text{joint}}(0, x, z_3, \dots, z_N) \\ &= \frac{1}{N(N-1)} \left(N K_N(x, x) - K_N(0, x)^2 \right). \end{aligned} \quad (\text{G.18})$$

En utilisant les résultats (G.8) et (G.9), la densité se réécrit en fonction des polynômes de Laguerre

$$\rho_{\max}(x, N) = \frac{N e^{-x}}{(N-1)x^2} \left[(Nx-1)L_{N-1}(x)^2 + (2-x)L_{N-1}(x)L_N(x) - x(N-1)L_N(x)L_{N-2}(x) - L_N(x)^2 \right]. \quad (\text{G.19})$$

À partir de ce résultat, on retrouve la loi de Marcenko-Pastur (9.7) dans le régime de volume

$$N\rho_{\max}(Nx, N) \xrightarrow{N \rightarrow \infty} \rho_{\text{MP}}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}, \quad (\text{G.20})$$

en effet, dans ce régime, les fluctuations de la valeur propre minimale sont imperceptibles, on peut donc remplacer λ_{\min} par sa moyenne 0 dans la définition de la densité à partir de la valeur propre minimale (G.17). Il est également possible d'étudier à partir de la formule (G.19) le régime de bord

— Régime de bord de la densité à partir du minimum —

$$\begin{aligned} \rho_{\max}\left(\frac{r}{N}, N\right) \xrightarrow{N \rightarrow \infty} \rho_E(r) &= f_K(r, r) - f_K(0, r)^2 \\ &= J_0(2\sqrt{r})^2 + J_1(2\sqrt{r})^2 \left(1 - \frac{1}{r}\right). \end{aligned} \quad (\text{G.21})$$

Les régimes asymptotiques sont alors donnés par

$$\rho_E(r) \sim \begin{cases} \frac{r^2}{12}, & r \rightarrow 0 \\ \frac{1}{\pi\sqrt{r}}, & r \rightarrow \infty, \end{cases} \quad (\text{G.22})$$

le comportement à petit r provenant des interactions entre les deux plus petites valeurs propres est encore une fois quadratique à cause du déterminant de Vandermonde présent dans la loi jointe (G.1) et le comportement à r grand se raccordant parfaitement avec la loi de Marcenko-Pastur $c = 1$ (9.7).

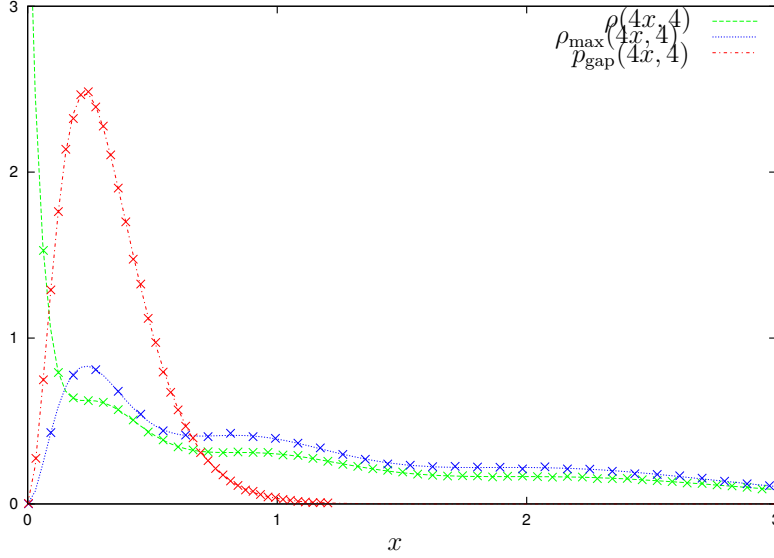


FIGURE G.1: Comparaisons entre les formules pour $N = 4$ de la densité (G.13), de la densité à partir du maximum (G.19) et du premier gap (G.25) et des résultats numériques obtenus en générant 10^6 matrices de Wishart de taille 4×4 .

G.4 Statistique du gap entre les deux plus petites valeurs propres

Il est possible de calculer la statistique du gap entre les deux plus petites valeurs propres en utilisant l'identité (G.2). Remarquons que dans ce cas, la formule (7.5) qui donne un lien entre la densité à partir du maximum et la statistique du premier gap n'est pas valable car, si l'on reprend la démonstration (7.3) et (7.4), le domaine des valeurs propres $\lambda < 0$ est interdit. En notant $\lambda_{i_{\min}} = \lambda_{\min}$ la plus petite valeur propre et λ_{i_2} la seconde, la statistique du premier gap est donnée par

$$p_{\text{gap}}(x, N) = \langle \delta(\lambda_{i_{\min}} - \lambda_{i_2} + x) \rangle. \quad (\text{G.23})$$

En utilisant l'identité (G.2), il est encore une fois possible d'exprimer cette quantité en termes du noyau (G.8) [113]

$$\begin{aligned} p_{\text{gap}}(x, N) &= N(N-1) \int_x^\infty d\lambda_{i_2} \int_{\lambda_{i_2}}^\infty d\lambda_3 \cdots \int_{\lambda_{i_2}}^\infty d\lambda_N P_{\text{joint}}(\lambda_{i_2} - x, \lambda_{i_2}, \lambda_3, \dots, \lambda_N) \\ &= N(N-1) \int_x^\infty d\lambda_{i_2} e^{-N\lambda_{i_2}} \int_0^\infty dz_3 \cdots \int_0^\infty dz_N P_{\text{joint}}(-x, 0, z_3, \dots, z_N) \\ &= (N-1)e^{-Nx} \int_0^\infty dz_3 \cdots \int_0^\infty dz_N P_{\text{joint}}(-x, 0, z_3, \dots, z_N) \\ &= \frac{e^{-Nx}}{N} \left(NK_N(-x, -x) - K_N(0, -x)^2 \right). \end{aligned} \quad (\text{G.24})$$

En utilisant les résultats (G.8) et (G.9), la statistique du premier gap se réécrit en fonction des polynômes de Laguerre

$$p_{\text{gap}}(x, N) = \frac{Ne^{-(N-1)x}}{x^2} \left[(2+x)L_{N-1}(x)L_N(-x) - (1+Nx)L_{N-1}(-x)^2 + (N-1)xL_{N-2}(-x)L_N(-x) - L_N(-x)^2 \right], \quad (\text{G.25})$$

dont on peut tirer le régime typique du premier gap

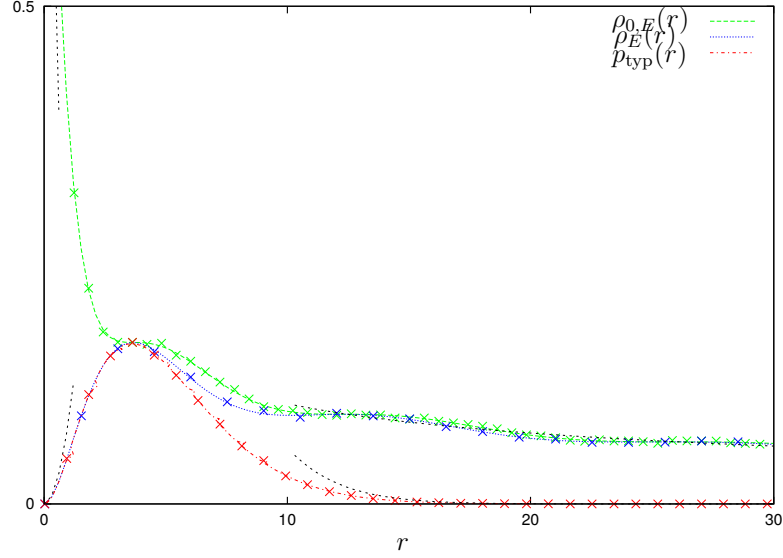


FIGURE G.2: Comparaisons entre les comportements de bord de la densité (G.15), de la densité à partir du maximum (G.21) et du premier gap (G.26) et des résultats numériques obtenus en générant 10^6 matrices de Wishart 50×50 . Les lignes pointillées représentent les différents comportements asymptotiques (G.16, G.22, G.28).

— Régime typique du premier gap —

$$\frac{1}{N} p_{\text{gap}} \left(\frac{r}{N}, N \right) \xrightarrow{N \rightarrow \infty} p_{\text{typ}}(r) = e^{-r} (f_K(-r, -r) - f_K(0, -r)^2) \quad (\text{G.26})$$

$$= e^{-r} \left(I_0(2\sqrt{r})^2 - I_1(2\sqrt{r})^2 \left(1 + \frac{1}{r} \right) \right), \quad (\text{G.27})$$

où I_0 et I_1 sont les deux premières fonctions de Bessel modifiées (16). Cette formule peut aussi être obtenue à partir du résultat (9.51) de Forrester et Hughes [66]. Les régimes asymptotiques de ce régime typique sont donnés par

$$p_{\text{typ}}(r) \sim \begin{cases} \frac{r^2}{12}, & r \rightarrow 0 \\ \frac{e^{4\sqrt{r}-r}}{8\pi r}, & r \rightarrow \infty, \end{cases} \quad (\text{G.28})$$

le comportement à petit r provenant encore une fois des interactions entre les deux plus petites valeurs propres (G.22). Le comportement à r grand correspond au premier ordre à la distribution de la valeur propre minimale (G.3). Nous traçons dans la figure G.1 les formules exactes pour la densité (G.13), la densité à partir du maximum (G.19) et le premier gap (G.25) pour $N = 4$ que nous comparons à des résultats numériques obtenus en générant 10^6 matrices de Wishart 4×4 . La figure G.2 reprend les comportements de bord de la densité (G.15), de la densité à partir du maximum (G.21) et du premier gap (G.26) comparer à des résultats numériques obtenus en générant 10^6 matrices de Wishart 50×50 .

Conclusion et perspectives

La statistique des valeurs extrêmes est une question majeure dans de nombreux contextes comme par exemple les sciences de l'environnement et la finance. En effet, dans ces différentes situations, les statistiques d'extrêmes ont des applications concrètes comme la prédiction de la distribution de probabilité des plus hautes crues de rivières, la taille des vagues scélérates, le montant d'importantes pertes financières dues aux fluctuations du marché boursier, etc... Récemment, l'étude des systèmes complexes et désordonnés a naturellement conduit les statistiques d'extrêmes à jouer un rôle crucial en physique statistique. En effet, la caractérisation des propriétés statistiques du maximum X_{\max} ou du minimum X_{\min} d'un ensemble de N variables aléatoires x_1, x_2, \dots, x_N a trouvé de nombreuses applications dans cette branche de la physique théorique. La description de la statistique d'un extremum global, X_{\max} ou X_{\min} , est certainement une caractéristique importante d'une collection de variables aléatoires mais celle-ci ne décrit les fluctuations que d'une seule variable parmi un grand nombre de variables aléatoires. Une question naturelle qui se pose alors est la suivante : ces valeurs extrêmes sont-elles isolées, loin des autres variables ou bien au contraire existe-t-il un grand nombre d'autres variables proches de ces valeurs extrêmes ? C'est une question cruciale dans le contexte des systèmes désordonnés où les propriétés de basse température sont gouvernées par des états excités proche du niveau d'énergie minimale (le niveau fondamental).

Ces questions ont naturellement débouché sur l'introduction de la densité d'état à partir de ces extrêmes (5) étudiée tout au long de cette thèse. Toutefois les résultats précédemment obtenus sur cette quantité étaient restreints à l'étude d'ensemble de variables aléatoires indépendantes et identiquement distribuées [139], cas où les statistiques d'extrêmes sont maintenant très bien comprises grâce notamment à l'identification des trois classes d'universalité : Gumbel, Fréchet et Weibull. Pourtant de nombreux modèles fondamentaux de physique statistique mettent en jeu de fortes corrélations entre les différents degrés de liberté du système. C'est le cas par exemple des marches au hasard (mouvement brownien), des matrices aléatoires, des modèles de croissance stochastique ou bien encore des modèles de systèmes étendus en milieu aléatoire (comme le polymère dirigé dans un potentiel désordonné). Il existe beaucoup moins de résultats dans les cas où les variables sont fortement corrélées qui possèdent une plus grande richesse de comportement. C'est dans ce but que s'inscrit cette thèse, afin d'étudier divers aspects de la théorie des extrêmes de variables aléatoires fortement corrélées sur les deux modèles que sont le mouvement brownien et les matrices aléatoires. Ces deux modèles sont non triviaux et permettent d'obtenir des résultats analytiques exacts.

La première partie de cette thèse concerne l'étude de la densité à partir du maximum pour différents mouvements browniens (brownien libre, pont, excursion, méandre, brownien réfléchi et pont réfléchi). J'ai calculé pour chacun des cas, en utilisant une méthode utilisant des propagateurs, la densité moyenne à partir du maximum (3.18), (3.25), (3.28), (3.29) et (3.30). Ces différentes distributions permettent d'affiner les modèles en fonction des contextes (ensemble de variables libres, périodiques, bornées...). J'ai également remar-

qué que la transformation de Vervaat, schématisée dans la figure 3.2, démontre que les statistiques d'extrêmes du pont et de l'excursion sont identiques. J'ai ensuite utilisé cette méthode pour étudier le cas de plusieurs marches aléatoires indépendantes ou contraintes à ne pas se croiser ainsi que l'étude temporelle des positions à une distance r du maximum dans le chapitre 4. Par la suite, j'ai développé dans le chapitre 5 une nouvelle approche théorique fondée sur la technique d'intégrale de chemin, qui permet d'étudier les fonctionnelles du maximum (encadré 5.22 dans le cas du brownien libre et 5.26 dans le cas du pont et de l'excursion). Cette nouvelle méthode m'a permis de calculer la distribution complète de la densité à partir du maximum pour ces cas (5.36) et (5.61). Enfin, j'ai appliqué cette méthode pour calculer la statistique d'une certaine famille de fonctionnelle du maximum (5.69) qui possède de multiples applications comme le calcul de l'aire sous un brownien, le coût de l'algorithme optimal de recherche du maximum d'une marche aléatoire ainsi que le temps maximal pour une particule balistique pour traverser un potentiel aléatoire brownien. Ces résultats suggèrent plusieurs extensions intéressantes, il est ainsi naturel de se demander ce que deviennent ces résultats pour des marches aléatoires en temps discrets ainsi que pour des marches de Lévy. Dans ces cas, il n'est pas possible d'utiliser la méthode l'intégrale de chemin, il est donc nécessaire de développer dans ces cas la méthode des propagateurs. Une façon différente d'étudier les statistiques des extrêmes et de considérer les statistiques d'ordre [21, 22, 115]. On s'intéresse alors non seulement à la statistique du (premier) maximum mais également aux propriétés du second, troisième, ..., $k^{\text{ième}}$ maximum. Ces statistiques d'ordres ont été étudiées pour des variables indépendantes ainsi que dans le cas d'une marche aléatoire en temps discret unidimensionnelle [142]. Il serait intéressant d'étudier ces quantités dans le cas du mouvement brownien ainsi que les connexions entre les statistiques des gaps et la densité à partir du maximum.

Dans la seconde partie, j'ai étudié le cas où la collection de variables aléatoires fortement corrélées est composée des N valeurs propres d'une matrice aléatoire $N \times N$. En premier lieu, nous avons utilisé dans le chapitre 7 une méthode fondée sur la méthode des polynômes orthogonaux pour calculer analytiquement la densité à partir du maximum (7.26) ainsi que la statistique de l'écart entre les deux plus grandes valeurs propres -le premier gap- (7.47) pour des matrices gaussiennes unitaires. J'ai obtenu une formule explicite pour le premier gap, différente de celle obtenue par Witte *et al.* [166], faisant intervenir des fonctions transcendentes de Painlevé. Il serait intéressant de montrer que ces deux formules coïncident. Je n'ai étudié dans ce cas que les fluctuations typiques du gap, il serait par ailleurs intéressant de poursuivre cette étude en étudiant les fluctuations dites atypiques (régime de grande déviation). J'ai ensuite généralisé dans la section 8.1 certains résultats, en particulier les régimes d'échelles et les comportements asymptotiques, au cas des matrices appartenant aux ensembles gaussiens β . La généralisation de ces résultats pour $\beta = 1$ permet de trouver de nouveaux régimes intéressants dans le modèle de verre de spin en champ moyen de Sherrington-Kirkpatrick. Il serait ainsi utile de calculer directement ces quantités pour $\beta = 1$. J'ai également obtenu, la densité à partir du maximum (8.27), ainsi que le premier gap (8.28), lorsque la valeur propre maximale est contrainte ce qui devrait permettre de mieux comprendre les statistiques d'extrêmes pour des modèles de matrices aléatoires dont la valeur propre maximum est fixée comme les matrices de Markov aléatoires. Il serait intéressant de poursuivre cette étude en analysant les différents gaps successifs : le second, le troisième, ... le $k^{\text{ième}}$ gap avec $k \ll \sqrt{N}$ pour rester dans le régime de bord, des résultats numériques préliminaires montrent en effet que cette quantité possède des lois d'échelle intéressantes. Enfin, comme nous l'avons mentionné (6.29), les positions de N fermions $1D$ sans spin confinés dans un potentiel harmonique sont distribuées suivant la même loi de probabilité jointe que les valeurs propres de l'ensemble gaussien unitaire [51, 111]. Il serait ainsi intéressant de

mesurer expérimentalement les propriétés de bord du spectre comme la densité à partir du maximum et le premier gap. Il serait également utile d'étudier l'évolution de ces quantités avec la température dans ce contexte [39]. J'ai également, dans le chapitre 9, adapté la méthode des polynômes orthogonaux pour étudier les matrices de Wishart. En utilisant cette méthode, j'ai redérivé les résultats sur les différentes statistiques de la valeur propre minimale obtenus par Tracy et Widom [155]. J'ai en effet montré que la distribution de la valeur propre minimale est gouvernée par une équation de Painlevé V lorsque le nombre de valeurs propres N est fini (9.38), par une équation de Painlevé III dans la limite dite de *hard edge* (9.50) et par une équation de Painlevé II, et donc la distribution de Tracy-Widom, dans la limite de *soft edge* (9.60). J'ai également prouvé une conjecture faite par Edelman *et al.* [47] sur la première correction de la distribution de la valeur propre minimale (9.55). À partir des développements (9.46), il serait intéressant de reproduire l'étude que l'on a fait dans le chapitre 7 pour étudier les statistiques de la densité à partir de la valeur propre minimale ainsi que le premier gap pour l'ensemble des matrices de Wishart dans la limite de *hard edge*.

Remerciements

Je termine ce manuscrit par les habituels remerciements. Je remercie tout d'abord les personnes qui lisent, ne serait-ce qu'en partie, cette thèse. Celle-ci est l'aboutissement de trois ans de travail et a été pour moi une expérience incroyablement riche.

En premier lieu, je souhaite remercier Malte Henkel et Gernot Akemann pour avoir accepté d'être rapporteurs de ma thèse. Je remercie plus particulièrement Gernot Akemann pour avoir organisé deux *workshops* sur les matrices aléatoires auxquels j'ai participé : Bielefeld 2013 et Brunel 2014. Ces deux séjours m'ont apporté énormément de connaissances et aussi de souvenirs. Je remercie Tom Claeys, Henk Hilhorst, Henri Orland et Clément Sire pour avoir accepté d'être membres du jury. Je ne remercierai jamais assez Henk pour son cours à l'ENS qui m'a donné l'envie de faire de la physique statistique (même si mon sujet s'est beaucoup éloigné de ce cours) mais surtout de m'avoir présenté Grégory pour encadrer mon stage de master.

Je me souviens encore de mon rendez-vous avec Grégory où il m'a présenté ce qui allait devenir mon sujet de stage. Ses explications claires ont immédiatement fait mouche. Gregory a été un super directeur. Sa manière de concevoir la science est pour moi un grand exemple et j'espère m'en être imprégné au maximum pendant ces années. Sa disponibilité, la liberté qu'il m'a accordée, sa bonne humeur et son encadrement ont fait de ces années de travail un vrai plaisir. Je suis fier d'avoir pu faire ma thèse avec lui.

Je remercie également le LPTMS, et plus généralement le campus de Paris-Sud. L'ambiance tranquille de ce labo a rendu ce travail encore plus plaisant. Je remercie spécialement Alain Comtet et Satya Majumdar pour leurs implications dans mon travail, le directeur du laboratoire Emmanuel Trizac ainsi que Claudine le Vaou, Géraldine Régis et Vincent Degat. Je remercie également mes camarades de sciences, thésards et postdocs, que j'ai pu côtoyer pendant ces trois ans.

J'adresse aussi mes remerciements au centre de préparation à l'agrégation qui, tout d'abord m'a permis d'obtenir ce concours, et où ensuite j'ai pu enseigner pendant ces trois ans. L'expérience que j'y ai accumulée me sera j'en suis sûr très utile. Un grand merci pour cela à tous ses membres et spécialement à Kenneth, JMR, Martial et Éric.

En dehors de ce cadre de recherche, il y a un très grand nombre de personnes que je veux remercier : mes différents colocataires qui m'auront supporté : Marine, Frederike & Félix, Jean-Paul, Clémence et Fabien. Je remercie également Élina & Amaury (& Aurèle) et Maxime ainsi que le cercle des profs : Sa Majesté John, Élise, Alex, Adrien (& les cinémas), Silvère, Romain, Manon & Sylvain (& Mimichat), Caroline & Michaël (& tous les enfants), Céline & Loic, Seb & Ana, Clément, Benoît, Jean-Romain, Jean-Christophe et Mathieu. Les différentes vacances prises avec vous m'ont donné de grandes bouffées d'air frais, de calme et de gaieté pour repartir de plus belle ! Je vous rejoins dès l'an prochain (mais pas géographiquement...) et j'espère que les prochaines vacances seront encore une fois parfaites. Je remercie également Pascal, Alan, Sylvain, Camille, Romain, Benjamin, Nathalie et les RecPc pour la bonne humeur qu'ils ont pu partager avec moi. Je remercie plus spécialement Jérôme pour son aide précieuse dans l'utilisation de LaTeX. Un grand

merci également aux personnes qui m'auront conforté dans l'idée de faire cette thèse : Stéphane Clemmen (et plus généralement les personnes que j'ai fréquentées à Bruxelles) et Jean François Allemand (qui, en plus de s'être si bien occupé de moi à l'ENS, m'aura poussé jusqu'au bout pour une thèse que je ne regrette pas).

Je conclus ce manuscrit en remerciant mes proches qui m'ont toujours soutenu, mes parents, mes frères et sœurs : Élodie & Baptiste, Joe & Alex, Lou, Elise, Emilie et Bryan. Bonne chance aux petits Dylan, Paul et Éloan. Promis, je suis de retour dans la Comté très bientôt.

Le monde est dangereux à vivre ! Non pas tant à cause de ceux qui font le mal, mais à cause de ceux qui regardent et laissent faire.

– Albert Einstein

#JeSuisCharlie

Bibliographie

- [1] AKEMANN, G. et ATKIN, M. R. (2013). Higher order analogues of Tracy–Widom distributions via the Lax method. *Journal of Physics A : Mathematical and Theoretical*, 46(1):015202.
- [2] AKEMANN, G., BAIK, J. et DI FRANCESCO, P. (2011). *The Oxford handbook of random matrix theory*. Oxford University Press.
- [3] ALVARADO, E., SANDBERG, D. V. et PICKFORD, S. G. (1998). *Modeling large forest fires as extreme events*. National Emergency Training Center.
- [4] ARNOLD, B. C., BALAKRISHNAN, N. et NAGARAJA, H. N. (1992). *A first course in order statistics*, volume 54. Siam.
- [5] BAIK, J., DEIFT, P. et JOHANSSON, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *Journal of the American Mathematical Society*, 12(4):1119–1178.
- [6] BAIK, J. et RAINS, E. M. (2000). Limiting distributions for a polynuclear growth model with external sources. *Journal of Statistical Physics*, 100(3-4):523–541.
- [7] BAŽANT, Z. P., PANG, S. D., VOŘECHOVSKÝ, M., NOVÁK, D. et PUKL, R. (2004). Statistical size effect in quasibrittle materials : computation and extreme value theory. *In Proc. of 5th Int. Conference FraMCoS–Fracture Mechanics of Concrete and Concrete Structures, Li, VC et al.(Eds.)*, volume 1, pages 189–196.
- [8] BELMEHDI, S. et RONVEAUX, A. (1994). Laguerre-Freud’s Equations for the Recurrence Coefficients of Semi-classical Orthogonal Polynomials. *Journal of Approximation Theory*, 76(3):351–368.
- [9] BENGTSSON, A. et NILSSON, C. (2007). Extreme value modelling of storm damage in Swedish forests. *Natural Hazards and Earth System Science*, 7(5):515–521.
- [10] BERG, H. C. (1983). *Random walks in biology*. Princeton University Press.
- [11] BERTALAN, Z., SHEKHAWAT, A., SETHNA, J. P. et ZAPPERI, S. (2014). Fracture Strength : Stress Concentration, Extreme Value Statistics, and the Fate of the Weibull Distribution. *Physical Review Applied*, 2(3):034008.
- [12] BERTIN, E. et CLUSEL, M. (2006). Generalized extreme value statistics and sum of correlated variables. *Journal of Physics A : Mathematical and General*, 39(24):7607.
- [13] BERTIN, E. et GYÖRGYI, G. (2010). Renormalization flow in extreme value statistics. *Journal of Statistical Mechanics : Theory and Experiment*, 2010(08):P08022.

- [14] BOUCHAUD, J.-P. et MÉZARD, M. (1997). Universality classes for extreme-value statistics. *Journal of Physics A : Mathematical and General*, 30(23):7997.
- [15] BOUCHAUD, J.-P. et POTTERS, M. (2003). *Theory of financial risk and derivative pricing : from statistical physics to risk management*. Cambridge university press.
- [16] BOWICK, M. J. et BRÉZIN, É. (1991). Universal scaling of the tail of the density of eigenvalues in random matrix models. *Physics Letters B*, 268(1):21–28.
- [17] BRAY, A. J. et WINKLER, K. (2004). Vicious walkers in a potential. *Journal of Physics A : Mathematical and General*, 37(21):5493.
- [18] BROWN, R. (1828). A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants ; and on the general existence of active molecules in organic and inorganic bodies. *Edinburgh Journal of Science*, 9(2,155):336–346.
- [19] BRUNET, É. et DERRIDA, B. (2000a). Ground state energy of a non-integer number of particles with δ attractive interactions. *Physica A : Statistical Mechanics and its Applications*, 279(1):398–407.
- [20] BRUNET, É. et DERRIDA, B. (2000b). Probability distribution of the free energy of a directed polymer in a random medium. *Physical Review E*, 61(6):6789.
- [21] BRUNET, E. et DERRIDA, B. (2009). Statistics at the tip of a branching random walk and the delay of traveling waves. *Europhysics Letters*, 87(6):60010.
- [22] BRUNET, É. et DERRIDA, B. (2011). A branching random walk seen from the tip. *Journal of Statistical Physics*, 143(3):420–446.
- [23] BUNDSCHUH, R. et HWA, T. (2000). An analytic study of the phase transition line in local sequence alignment with gaps. *Discrete Applied Mathematics*, 104(1):113–142.
- [24] BURKHARDT, T. W., GYÖRGYI, G., MOLONEY, N. R. et RÁCZ, Z. (2007). Extreme statistics for time series : Distribution of the maximum relative to the initial value. *Physical Review E*, 76(4):041119.
- [25] CAVALLI-SFORZA, L. L., MENOZZI, P. et PIAZZA, A. (1994). *The history and geography of human genes*. Princeton university press.
- [26] CHASSAING, P. (1999). How many probes are needed to compute the maximum of a random walk? *Stochastic processes and their applications*, 81(1):129–153.
- [27] CHASSAING, P. et LOUCHARD, G. (2002). Reflected Brownian bridge area conditioned on its local time at the origin. *Journal of algorithms*, 44(1):29–51.
- [28] CHASSAING, P., MARCKERT, J. F. et YOR, M. (2003). A stochastically quasi-optimal search algorithm for the maximum of the simple random walk. *The Annals of Applied Probability*, 13(4):1264–1295.
- [29] CHOWDHURY, D. (2005). 100 years of Einstein’s theory of Brownian motion : from pollen grains to protein trains-1. *Resonance-Journal of Science Education*, 10(9):63–78.
- [30] CHUNG, K. L. (1976). Excursions in Brownian motion. *Arkiv för matematik*, 14(1):155–177.

- [31] CLAEYS, T., KRASOVSKY, I. et ITS, A. (2010). Higher-order analogues of the Tracy-Widom distribution and the Painlevé ii hierarchy. *Communications on pure and applied mathematics*, 63(3):362–412.
- [32] CUGLIANDOLO, L. F. et DEAN, D. S. (1995). Full dynamical solution for a spherical spin-glass model. *Journal of Physics A : Mathematical and General*, 28(15):4213.
- [33] DANIELS, H. E. (1945). The statistical theory of the strength of bundles of threads. i. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 183(995):405–435.
- [34] DARLING, D. A. (1983). On the supremum of a certain Gaussian process. *The Annals of Probability*, 11(3):803–806.
- [35] DAVID, H. A. et NAGARAJA, H. N. (1970). *Order statistics*. Wiley Online Library.
- [36] de GENNES, P.-G. (1968). Soluble model for fibrous structures with steric constraints. *The Journal of Chemical Physics*, 48(5):2257–2259.
- [37] de HAAN, L. et RESNICK, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of Probability*, pages 97–124.
- [38] de HAAN, L. et STADTMÜLLER, U. (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)*, 61(03):381–395.
- [39] DEAN, D. S., LE DOUSSAL, P., MAJUMDAR, S. N. et SCHEHR, G. (2014). Finite temperature free fermions and the Kardar-Parisi-Zhang equation at finite time. *arXiv preprint 1412.1590*.
- [40] DEAN, D. S. et MAJUMDAR, S. N. (2001). Extreme-value statistics of hierarchically correlated variables deviation from Gumbel statistics and anomalous persistence. *Physical Review E*, 64(4):046121.
- [41] DEAN, D. S. et MAJUMDAR, S. N. (2006). Large deviations of extreme eigenvalues of random matrices. *Physical review letters*, 97(16):160201.
- [42] DEIFT, P. et GIOEV, D. (2007). Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. *Communications on pure and applied mathematics*, 60(6):867–910.
- [43] DOMÍNGUEZ, M. R. et ARGANIS, J. M. (2012). Validation of methods to estimate design discharge flow rates for dam spillways with large regulating capacity. *Hydrological Sciences Journal*, 57(3):460–478.
- [44] DUMITRIU, I. et EDELMAN, A. (2002). Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847.
- [45] DUPLANTIER, B. (2006). Brownian Motion, “Diverse and Undulating”. In *Einstein, 1905-2005 : Poincaré Seminar 2005*, volume 47, page 201. Springer Science & Business Media.
- [46] DYSON, F. J. (1962). A brownian-motion model for the eigenvalues of a random matrix. *Journal of Mathematical Physics*, 3(6):1191–1198.

- [47] EDELMAN, A., GUIONNET, A. et PÉCHÉ, S. (2014). Beyond Universality in Random Matrix Theory. *arXiv preprint 1405.7590*.
- [48] EICHNER, J. F., KANTELHARDT, J. W., BUNDE, A. et HAVLIN, S. (2006). Extreme value statistics in records with long-term persistence. *Physical Review E*, 73(1):016130.
- [49] EINMAHL, J. H. J. et MAGNUS, J. R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association*, 103(484).
- [50] EINSTEIN, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der physik*, 322(8):549–560.
- [51] EISLER, V. et RÁ CZ, Z. (2013). Full counting statistics in a propagating quantum front and random matrix spectra. *Physical review letters*, 110(6):060602.
- [52] ELKIES, N., KUPERBERG, G., LARSEN, M. et PROPP, J. (1992). Alternating-sign matrices and domino tilings (part i & ii). *Journal of Algebraic Combinatorics*, 1(2 & 3):111–132 & 219–234.
- [53] EMBRECHTS, P., KLÜPPELBERG, C. et MIKOSCH, T. (1997). *Modelling extremal events*, volume 33. Springer Science & Business Media.
- [54] ESSAM, J. W. et GUTTMANN, A. (1995). Vicious walkers and directed polymer networks in general dimensions. *Physical Review E*, 52(6):5849.
- [55] FAHMI, K. J. et AL ABBASI, J. N. (1987). Mixture distributions-an alternative approach for estimating maximum magnitude earthquake occurrence. *Geophysical Journal International*, 89(2):741–747.
- [56] FELLER, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, pages 427–432.
- [57] FELLER, W. (1957). An introduction to probability theory and its applications, vol. 1&2.
- [58] FISHER, D. S. et HUSE, D. A. (1988a). Equilibrium behavior of the spin-glass ordered phase. *Physical Review B*, 38(1):386.
- [59] FISHER, D. S. et HUSE, D. A. (1988b). Nonequilibrium dynamics of spin glasses. *Physical Review B*, 38(1):373.
- [60] FISHER, M. E. (1984). Walks, walls, wetting, and melting. *Journal of Statistical Physics*, 34(5-6):667–729.
- [61] FISHER, R. A. et TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.
- [62] FLAJOLET, P., POBLETE, P. et VIOLA, A. (1999). On the analysis of linear probing hashing. *Algorithmica*, 22(4):490–515.
- [63] FOKAS, A., ITS, A., KAPAEV, A. et NOVOKSHENOV, V. (2006). Painlevé Transcendents : The Riemann-Hilbert Approach.

- [64] FORRESTER, P. J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402(3):709–728.
- [65] FORRESTER, P. J. (2010). *Log-gases and random matrices (LMS-34)*. Princeton University Press.
- [66] FORRESTER, P. J. et HUGHES, T. D. (1994). Complex Wishart matrices and conductance in mesoscopic systems : exact results. *Journal of Mathematical Physics*, 35(12):6736–6747.
- [67] FORRESTER, P. J., MAJUMDAR, S. N. et SCHEHR, G. (2011). Non-intersecting Brownian walkers and Yang–Mills theory on the sphere. *Nuclear Physics B*, 844(3):500–526.
- [68] FORRESTER, P. J. et WITTE, N. S. (2007). The distribution of the first eigenvalue spacing at the hard edge of the Laguerre unitary ensemble. *Kyushu Journal of Mathematics*, 61(2):457–526.
- [69] FORSTER, D., NELSON, D. R. et STEPHEN, M. J. (1977). Large-distance and long-time properties of a randomly stirred fluid. *Physical Review A*, 16(2):732.
- [70] FORTIN, J.-Y. et CLUSEL, M. (2015). Applications of extreme value statistics in physics. *Journal of Physics A : Mathematical and Theoretical*, 48(18):183001.
- [71] FRÉCHET, M. (1927). Sur la loi de probabilité de l'écart maximum. In *Annales de la société Polonaise de Mathématique*, volume 6, pages 93–116.
- [72] FREY, E. et KROY, K. (2005). Brownian motion : a paradigm of soft matter and biological physics. *Annalen der Physik*, 14(1-3):20–50.
- [73] FUKUNAGA, K. (2013). *Introduction to statistical pattern recognition*. Academic press.
- [74] FYODOROV, Y. V. (2005). Introduction to the random matrix theory : Gaussian unitary ensemble and beyond. *London Mathematical Society Lecture Note Series*, 322: 31.
- [75] GITTENBERGER, B. et LOUCHARD, G. (1999). The Brownian excursion multi-dimensional local time density. *Journal of applied probability*, 36(2):350–373.
- [76] GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453.
- [77] GUMBEL, E. (1958). Statistics of extremes. 1958. *Columbia Univ. press, New York*.
- [78] GUMBEL, E. J. (1935). Les valeurs extrêmes des distributions statistiques. In *Annales de l'institut Henri Poincaré*, volume 5, pages 115–158. Presses universitaires de France.
- [79] GYÖRGYI, G., MOLONEY, N. R., OZOGÁNY, K. et RÁCZ, Z. (2008). Finite-size scaling in extreme statistics. *Physical review letters*, 100(21):210601.
- [80] GYÖRGYI, G., MOLONEY, N. R., OZOGÁNY, K., RÁCZ, Z. et DROZ, M. (2010). Renormalization-group theory for finite-size scaling in extreme statistics. *Physical Review E*, 81(4):041135.
- [81] HENKEL, M., PLEIMLING, M., GODRÈCHE, C. et LUCK, J.-M. (2001). Aging, phase ordering, and conformal invariance. *Physical review letters*, 87(26):265701.

- [82] HOOGHIEMSTRA, G. (1982). On the explicit form of the density of Brownian excursion local time. *Proceedings of the American Mathematical Society*, 84(1):127–130.
- [83] HUSE, D. A. et HENLEY, C. L. (1985). Pinning and roughening of domain walls in Ising systems due to random impurities. *Physical review letters*, 54(25):2708.
- [84] HWA, T. et LÄSSIG, M. (1996). Similarity detection and localization. *Physical review letters*, 76(14):2591.
- [85] HWANG, H.-K. (1997). A constant arising from the analysis of algorithms for determining the maximum of a random walk. *Random Structures and Algorithms*, 10(3):333–335.
- [86] IMBRIE, J. Z. et SPENCER, T. (1988). Diffusion of directed polymers in a random environment. *Journal of Statistical Physics*, 52(3-4):609–626.
- [87] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, pages 475–501.
- [88] JANSON, S. et LOUCHARD, G. (2007). Tail estimates for the Brownian excursion area and other Brownian areas. *Electronic Journal of Probability [electronic only]*, 12:1600–1632.
- [89] JOCKUSH, W., PROPP, J. et SHOR, P. (1995). Random domino tilings and the arctic circle theorem, preprint, 1995, xxx. lanl. gov/abs. *arXiv preprint math.CO/9801068*.
- [90] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476.
- [91] JOHANSSON, K. (2002). Non-intersecting paths, random tilings and random matrices. *Probability theory and related fields*, 123(2):225–280.
- [92] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327.
- [93] KAHANE, J.-P. (2006). Le mouvement brownien et son histoire, réponses à quelques questions. *Images des Mathématiques, CNRS*.
- [94] KARDAR, M., PARISI, G. et ZHANG, Y.-C. (1986). Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889.
- [95] KATZ, R. W., PARLANGE, M. B. et NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8):1287–1304.
- [96] KOLMOGOROV, A. N. (1933). *Sulla determinazione empirica di una legge di distribuzione*. na.
- [97] KRIVINE, H. (2003). *Exercices de mathématiques pour physiciens, corrigés et commentés*. Cassini.
- [98] KURCHAN, J. et LALOUX, L. (1996). Phase space geometry and slow dynamics. *Journal of Physics A : Mathematical and General*, 29(9):1929.
- [99] LE DOUSSAL, P. et MONTHUS, C. (2003). Exact solutions for the statistics of extrema of some random 1D landscapes, application to the equilibrium and the dynamics of the toy model. *Physica A : Statistical Mechanics and its Applications*, 317(1):140–198.

- [100] LOUCHARD, G. (1984). Kac's formula, Levy's local time and Brownian excursion. *Journal of Applied Probability*, pages 479–499.
- [101] LÉVY, P. (1940). Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 7:283–339.
- [102] MAGNUS, A. P. (1995). Painlevé-type differential equations for the recurrence coefficients of semi-classical orthogonal polynomials. *Journal of computational and applied mathematics*, 57(1):215–237.
- [103] MAJUMDAR, S. N. (2005). Brownian Functionals in Physics and Computer Science. *Current Science*, 89:2076.
- [104] MAJUMDAR, S. N. et BOUCHAUD, J.-P. (2008). Optimal time to sell a stock in the Black–Scholes model : comment on ‘Thou Shalt Buy and Hold’, by A. Shiryaev, Z. Xu and XY Zhou. *Quantitative Finance*, 8(8):753–760.
- [105] MAJUMDAR, S. N. et COMTET, A. (2004). Exact maximal height distribution of fluctuating interfaces. *Physical review letters*, 92(22):225501.
- [106] MAJUMDAR, S. N. et COMTET, A. (2005). Airy distribution function : from the area under a Brownian excursion to the maximal height of fluctuating interfaces. *Journal of statistical physics*, 119(3-4):777–826.
- [107] MAJUMDAR, S. N., MOUNAIX, P. et SCHEHR, G. (2013a). Exact statistics of the gap and time interval between the first two maxima of random walks and Lévy flights. *Physical review letters*, 111(7):070601.
- [108] MAJUMDAR, S. N. et SCHEHR, G. (2014). Top eigenvalue of a random matrix : large deviations and third order phase transition. *Journal of Statistical Mechanics : Theory and Experiment*, 2014(1):P01012.
- [109] MAJUMDAR, S. N., SCHEHR, G., VILLAMAINA, D. et VIVO, P. (2013b). Large deviations of the top eigenvalue of large Cauchy random matrices. *Journal of Physics A : Mathematical and Theoretical*, 46(2):022001.
- [110] MAJUMDAR, S. N. et VERGASSOLA, M. (2009). Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices. *Physical review letters*, 102(6):060601.
- [111] MARINO, R., MAJUMDAR, S. N., SCHEHR, G. et VIVO, P. (2014). Phase transitions and edge scaling of number variance in Gaussian random matrices. *Physical review letters*, 112(25):254101.
- [112] MAY, R. M. (1972). Will a large complex system be stable? *Nature*, 238:413–414.
- [113] MEHTA, M. L. (2004). Random matrices.
- [114] MÉZARD, M. (1990). On the glassy nature of random directed polymers in two dimensions. *Journal de Physique*, 51(17):1831–1846.
- [115] MOLONEY, N. R., OZOGÁNY, K. et RÁCZ, Z. (2011). Order statistics of $1/f^\alpha$ signals. *Physical Review E*, 84(6):061101.

- [116] MONTHUS, C. et LE DOUSSAL, P. (2004). Low-temperature properties of some disordered systems from the statistical properties of nearly degenerate two-level excitations. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(4):535–548.
- [117] NADAL, C. (2011). *Matrices aléatoires et leurs applications à la physique statistique et quantique*. Thèse de doctorat, Paris 11.
- [118] NADAL, C. et MAJUMDAR, S. N. (2011). A simple derivation of the Tracy–Widom distribution of the maximal eigenvalue of a Gaussian unitary random matrix. *Journal of Statistical Mechanics : Theory and Experiment*, 2011(04):P04001.
- [119] ODLYZKO, A. M. (1995). Search for the maximum of a random walk. *Random Structures & Algorithms*, 6(2-3):275–295.
- [120] OMORI, F. (1894). On the after-shocks of earthquakes. 7(111).
- [121] PARK, S.-C., SZENDRO, I. G., NEIDHART, J. et KRUG, J. (2014). Phase transition in random adaptive walks on correlated fitness landscapes. *arXiv preprint 1408.4856*.
- [122] PERRET, A., COMTET, A., MAJUMDAR, S. N. et SCHEHR, G. (2013). Near-extreme statistics of Brownian motion. *Physical review letters*, 111(24):240601.
- [123] PERRET, A., COMTET, A., MAJUMDAR, S. N. et SCHEHR, G. (2015a). On certain functionals of the maximum of Brownian motion and their applications. *arXiv preprint 1502.01218 soumis à Journal of Statistical Physics*.
- [124] PERRET, A., FYODOROV, Y. V. et SCHEHR, G. (2015b). Notes on near extreme eigenvalues of GOE and the late time dynamics of the spherical SK models. *à paraître*.
- [125] PERRET, A., RISTIVOJEVIC, Z., LE DOUSSAL, P., SCHEHR, G. et WIESE, K. J. (2012). Super-Rough Glassy Phase of the Random Field X Y Model in Two Dimensions. *Physical review letters*, 109(15):157205.
- [126] PERRET, A. et SCHEHR, G. (2014). Near-extreme eigenvalues and the first gap of Hermitian random matrices. *Journal of Statistical Physics*, 156(5):843–876.
- [127] PERRET, A. et SCHEHR, G. (2015a). The density of eigenvalues seen from the soft edge of random matrices in the Gaussian β -ensembles. *arXiv preprint 1506.00245 pour Random Matrix Theory : Foundations and Applications, Kraków*.
- [128] PERRET, A. et SCHEHR, G. (2015b). Finite N corrections to the distribution of the smallest eigenvalue of Wishart-Laguerre complex matrices. *arXiv preprint 1506.02387 soumis à Random Matrices : Theory and Applications*.
- [129] PETERSEN, A. M., WANG, F., HAVLIN, S. et STANLEY, H. E. (2010). Market dynamics immediately before and after financial shocks : Quantifying the Omori, productivity, and Bath laws. *Physical Review E*, 82(3):036114.
- [130] PRÄHOFFER, M. et SPOHN, H. (2000a). Statistical self-similarity of one-dimensional growth processes. *Physica A : Statistical Mechanics and its Applications*, 279(1):342–352.
- [131] PRÄHOFFER, M. et SPOHN, H. (2000b). Universal distributions for growth processes in $1+1$ dimensions and random matrices. *Physical review letters*, 84(21):4882.

- [132] PRÄHOFFER, M. et SPOHN, H. (2002). Scale invariance of the PNG droplet and the Airy process. *Journal of statistical physics*, 108(5-6):1071–1106.
- [133] PRÄHOFFER, M. et SPOHN, H. (2004). Exact scaling functions for one-dimensional stationary KPZ growth. *Journal of statistical physics*, 115(1-2):255–279.
- [134] PREISENDORFER, R. W. et MOBLEY, C. D. (1988). *Principal component analysis in meteorology and oceanography*, volume 17. Elsevier Amsterdam.
- [135] PROPP, J. (2003). Generalized domino-shuffling. *Theoretical Computer Science*, 303(2):267–301.
- [136] RAMBEAU, J. (2011). *Statistiques d’extrêmes d’interfaces en croissance*. Thèse de doctorat, Paris 11.
- [137] RAMBEAU, J. et SCHEHR, G. (2011). Distribution of the time at which n vicious walkers reach their maximal height. *Physical Review E*, 83(6):061146.
- [138] RANDON-FURLING, J. (2009). *Statistiques d’extrêmes du mouvement brownien et applications*. Thèse de doctorat, Université Paris Sud-Paris XI.
- [139] SABHAPANDIT, S. et MAJUMDAR, S. N. (2007). Density of near-extreme events. *Physical review letters*, 98(14):140201.
- [140] SCHEHR, G. (2011). *Des systèmes élastiques désordonnés aux statistiques d’événements rares*. Habilitation à diriger des recherches, Université Paris Sud-Paris XI.
- [141] SCHEHR, G. (2012). Extremes of N vicious walkers for large N : application to the directed polymer and KPZ interfaces. *Journal of Statistical Physics*, 149(3):385–410.
- [142] SCHEHR, G. et MAJUMDAR, S. N. (2012). Universal order statistics of random walks. *Physical review letters*, 108(4):040601.
- [143] SCHEHR, G. et MAJUMDAR, S. N. (2014). Exact record and order statistics of random walks via first-passage ideas, to appear in the special volume” First-Passage Phenomena and Their Applications”, eds. *R. Metzler, G. Oshanin, S. Redner. World Scientific*.
- [144] SCHEHR, G., MAJUMDAR, S. N., COMTET, A. et RANDON-FURLING, J. (2008). Exact distribution of the maximal height of p vicious walkers. *Physical review letters*, 101(15):150601.
- [145] SHURYAK, E. V. et VERBAARSCHOT, J. (1993). Random matrix theory and spectral sum rules for the Dirac operator in QCD. *Nuclear Physics A*, 560(1):306–320.
- [146] SMITH, L. I. (2002). A tutorial on principal components analysis. *Cornell University, USA*, 51:52.
- [147] SZEGÖ, G. (1975). *Orthogonal polynomials*, volume 4th edition. American Mathematical Soc. : Providence R.I.
- [148] TAKÁCS, L. (1991). A Bernoulli excursion and its various applications. *Advances in Applied Probability*, pages 557–585.
- [149] TAKÁCS, L. (1992). Limit theorems for random trees. *Proceedings of the National Academy of Sciences*, 89(11):5011–5014.

- [150] TAKÁCS, L. (1995a). Brownian local times. *International Journal of Stochastic Analysis*, 8(3):209–232.
- [151] TAKÁCS, L. (1995b). Limit distributions for the Bernoulli meander. *Journal of applied probability*, pages 375–395.
- [152] TAKEUCHI, K. A. et SANO, M. (2010). Universal fluctuations of growing interfaces : evidence in turbulent liquid crystals. *Physical review letters*, 104(23):230601.
- [153] TRACY, C. A. et WIDOM, H. (1993). Introduction to random matrices. *In Geometric and quantum aspects of integrable systems*, pages 103–130. Springer.
- [154] TRACY, C. A. et WIDOM, H. (1994a). Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159(1):151–174.
- [155] TRACY, C. A. et WIDOM, H. (1994b). Level spacing distributions and the Bessel kernel. *Communications in mathematical physics*, 161(2):289–309.
- [156] TRACY, C. A. et WIDOM, H. (2007). Nonintersecting brownian excursions. *The Annals of Applied Probability*, 17(3):953–979.
- [157] UVAROV, V. (1959). Relation between polynomials orthogonal with different weights. *Doklady Akademii Nauk SSSR*, 126(1):33 – 36.
- [158] UVAROV, V. (1969). The connection between systems of polynomials orthogonal with respect to different distribution functions. *{USSR} Computational Mathematics and Mathematical Physics*, 9(6):25 – 36.
- [159] VAN KAMPEN, N. G. (2007). Stochastic processes in physics and chemistry, (north-holland personal library).
- [160] VERE-JONES, D. (1969). A note on the statistical interpretation of Båth’s law. *Bulletin of the Seismological Society of America*, 59(4):1535–1541.
- [161] VERVAAT, W. (1979). A relation between Brownian bridge and Brownian excursion. *The Annals of Probability*, pages 143–149.
- [162] VON SMOLUCHOWSKI, M. (1906). Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, 326(14):756–780.
- [163] WIGNER, E. P. (1951). On the statistical distribution of the widths and spacings of nuclear resonance levels. *In Mathematical Proceedings of the Cambridge Philosophical Society*, volume 47, pages 790–798. Cambridge Univ Press.
- [164] WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, pages 325–327.
- [165] WISHART, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52.
- [166] WITTE, N. S., BORNEMANN, F. et FORRESTER, P. (2013). Joint distribution of the first and second eigenvalues at the soft edge of unitary ensembles. *Nonlinearity*, 26(6):1799.

Publications

Near-extreme statistics of brownian motion.

[122] A. Perret, A. Comtet, S.N. Majumdar, G. Schehr (2013)
Physical review letters, 111(24) :240601.

Near-Extreme Statistics of Brownian Motion

Anthony Perret,^{1,*} Alain Comtet,^{1,2,†} Satya N. Majumdar,^{1,‡} and Grégory Schehr^{1,§}¹Université Paris-Sud-Paris 11, CNRS, LPTMS, 91405 Orsay Cedex, France²Université Pierre et Marie Curie-Paris 6, 75252 Paris Cedex 05, France

(Received 30 May 2013; revised manuscript received 18 September 2013; published 9 December 2013)

We study the statistics of near-extreme events of Brownian motion (BM) on the time interval $[0, t]$. We focus on the density of states near the maximum, $\rho(r, t)$, which is the amount of time spent by the process at a distance r from the maximum. We develop a path integral approach to study functionals of the maximum of BM, which allows us to study the full probability density function of $\rho(r, t)$ and obtain an explicit expression for the moments $\langle [\rho(r, t)]^k \rangle$ for arbitrary integer k . We also study near extremes of constrained BM, like the Brownian bridge. Finally we also present numerical simulations to check our analytical results.

DOI: 10.1103/PhysRevLett.111.240601

PACS numbers: 05.40.Fb, 02.50.Cw, 05.40.Jc

Introduction.—Since its first developments in the early 1930s, extreme value statistics (EVS) have found an increasing number of applications. Besides the fields of engineering [1], natural sciences [2], or finance [3,4], where EVS have been applied for a long time, extreme value questions play also now a key role in physics [5–7].

The standard question of EVS concerns the maximum X_{\max} (or the minimum X_{\min}) among a collection of N random variables X_1, \dots, X_N . However the fluctuations of this global quantity X_{\max} give only a partial information about the extreme events in this sequence of random variables. For instance, if X_i 's represent the energy levels of a disordered system, the low but finite temperature physics of this system is instead determined by the statistical properties of the states with an energy close to the ground state, i.e., “near minimal” states [7–10]. Near-extreme events are naturally related to the subject of order statistics [11], where one considers not only the first maximum X_{\max} but also the second or third one, and more generally the k th maximum. Order statistics recently arose in various problems of statistical physics to characterize the crowding of near extremes [12–16].

Besides their relevance in physics, near extremes are also important for various applied sciences. This is for instance the case in natural sciences or in finance where extreme events like earthquakes or financial crashes are usually preceded and followed by foreshocks and aftershocks [17–19]. This is also a natural question in climatology where a maximal (or minimal) temperature is usually accompanied by a heat (or cold) wave, which can have drastic consequences [20,21]. Similar questions arise in the context of sporting events, like marathon packs [22].

In all these situations a natural and useful quantity to characterize the crowding of near extremes is the density of states (DOS) near the maximum, $\rho(r, t)$ [20]. For a continuous stochastic process $x(\tau)$ in the time interval $[0, t]$, the DOS is defined as

$$\rho(r, t) = \int_0^t \delta[x_{\max} - x(\tau) - r] d\tau, \quad (1)$$

where $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$. Hence, $\rho(r, t) dr$ denotes the amount of time spent by $x(\tau)$ at a distance within the interval $[r, r + dr]$ from x_{\max} (see Fig. 1). Hence $\rho(r, t)$ is similar to the so-called “local time” [23] $T_{\text{loc}}(r, t) = \int_0^t \delta[x(\tau) - r] d\tau$, with the major difference that in Eq. (1) the distances are measured from x_{\max} , which is itself a random variable. Note that, by definition, $\int_0^\infty \rho(r, t) dr = t$. Clearly, $\rho(r, t)$ is a random variable as it fluctuates from one realization of $\{x(\tau)\}_{0 \leq \tau \leq t}$ to another one: an important question is then to characterize its fluctuations.

This question has attracted much attention during the last 15 years, both in statistics [24,25], often motivated by problems related to insurance risks, and more recently in statistical physics [20], and in econophysics [26]. Despite important literature on this subject, the only available

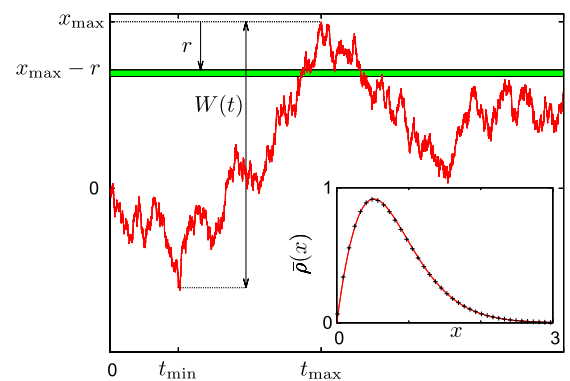


FIG. 1 (color online). One realization of the stochastic process $x(\tau)$ on the time interval $[0, t]$, with a width $W(t) = \max_{0 \leq \tau \leq t} x(\tau) - \min_{0 \leq \tau \leq t} x(\tau)$. $x(\tau)$ spends a time $\rho(r, t) dr$ at a distance within $[r, r + dr]$ (the green stripe) from the maximum x_{\max} , with $\rho(r, t)$ being the DOS (1). Inset: The average DOS for BM $\langle \rho(r, t) \rangle = \sqrt{t} \bar{p}(r/\sqrt{t})$ where the exact scaling function $\bar{p}(x)$ in Eq. (2) is compared to simulations.

results concern independent and identically distributed random variables, where $x(\tau_1)$ and $x(\tau_2)$ for $\tau_1 \neq \tau_2$ are uncorrelated (or weakly correlated [20]). Yet, many situations where near extremes are important, like disordered systems or earthquakes statistics, involve strongly correlated variables. Recent studies in physics, like the fluctuations at the tip of the branching Brownian motion [12,13], or order statistics of time series displaying $1/f^\alpha$ correlations [14], including Brownian motion (BM) [15,16], have also unveiled the importance of near-extreme statistics for strongly correlated variables. Hence, any exact result on near extremes of strongly correlated variables would be of wide interest.

In this Letter, we make a first step in that direction and focus on the case where $x(\tau)$ is a one-dimensional Brownian motion. It starts from $x(0) = 0$, and evolves via $\dot{x}(\tau) = \zeta(\tau)$, $\zeta(\tau)$ being Gaussian white noise, $\langle \zeta(\tau)\zeta(\tau') \rangle = \delta(\tau - \tau')$. In this case, the time series $\{x(\tau)\}_{0 \leq \tau \leq t}$ is clearly a set of strongly correlated variables as $\langle x(\tau_1)x(\tau_2) \rangle = \min(\tau_1, \tau_2)$ [and $\langle x(\tau_1) \rangle = \langle x(\tau_2) \rangle = 0$]. For this simple yet nontrivial strongly correlated process, we are able to provide a complete analytical characterization of the statistics of $\rho(r, t)$. Let us begin by summarizing our main results.

We first focus on the average DOS and show that $\langle \rho(r, t) \rangle = t^{1/2} \bar{\rho}(r/t^{1/2})$, such that $\int_0^\infty \langle \rho(r, t) \rangle dr = t$, with

$$\begin{aligned} \bar{\rho}(x) &= 8[h(x) - h(2x)], \\ h(x) &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} - \frac{x}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right), \end{aligned} \quad (2)$$

where $\operatorname{erfc}(x) = (2/\sqrt{\pi}) \int_x^\infty e^{-y^2} dy$. In the inset of Fig. 1 we show a plot of $\bar{\rho}(x)$. It behaves linearly, $\bar{\rho}(x) \sim 4x$ for $x \rightarrow 0$, vanishes as $\bar{\rho}(x) \propto x^{-2} e^{-x^2/2}$ for $x \rightarrow \infty$, and exhibits a maximum for $x_{\text{typ}} = 0.51454\dots$, slightly smaller than the average value $x_{\text{ave}} = \sqrt{2/\pi} = 0.79788\dots$. The fact that $\bar{\rho}(x)$ does not vanish too rapidly as $x \rightarrow 0$ indicates that, on average, there is no gap between x_{max} and the rest of the crowd; hence “ x_{max} is not lonely at the top.” The mean DOS for BM (2) is thus quite different from the independent and identically distributed case [20]; in that case, depending on whether the tail of the parent distribution of the X_i 's decays slower than, faster than, or as a pure exponential, the limiting mean DOS converges to three different limiting forms, which are clearly different from Eq. (2).

The DOS is a random variable (1) and its average is not sufficient to characterize its statistics. We thus study its full probability density function (PDF) $P_t(\rho, r)$, as a function of ρ , for different values of the parameter r . This PDF is a particular case of a functional of the maximum of the BM. In this Letter, we establish a general framework, using path integral, to study such functionals of x_{max} and obtain $P_t(\rho, r)$ exactly. We show that it has an unusual form with a peak $\propto \delta(\rho)$ at $\rho = 0$, in addition to a nontrivial

continuous background density $p_t(\rho, r)$ for $\rho > 0$. We show that the amplitude of this peak $\propto \delta(\rho)$ has a probabilistic interpretation, so that $P_t(\rho, r)$ reads

$$P_t(\rho, r) = F_W(r, t) \delta(\rho) + p_t(\rho, r), \quad (3)$$

where $F_W(r, t) = \text{Prob.}[W(t) \leq r]$, given in Eq. (13), is the probability that the width $W(t) = \max_{0 \leq \tau \leq t} x(\tau) - \min_{0 \leq \tau \leq t} x(\tau)$ is smaller than r . This can be understood because if $W(t)$ is smaller than r , the amount of time spent by the process at a distance within $[r, r + dr]$ from the maximum is 0 (see Fig. 1), yielding the delta peak at $\rho = 0$. On the other hand, in Eq. (3), $p_t(\rho, r) = t^{-(1/2)} p_1(\rho/\sqrt{t}, r/\sqrt{t})$ is a regular function of ρ , for $r > 0$ (see Fig. 2). We obtain an explicit expression of its Laplace transform (LT) with respect to t given below [Eq. (12)]. From it we extract the asymptotic behaviors

$$p_1(\rho, r) = \begin{cases} p_1(0, r) + \mathcal{O}(\rho), & \rho \rightarrow 0 \\ \frac{\rho^2}{\sqrt{2\pi}} e^{-(\rho+2r)^2/2} [1 + \mathcal{O}(\rho^{-1})], & \rho \rightarrow \infty \end{cases} \quad (4)$$

where $p_1(0, r)$ is a nontrivial function of r , given in Eq. (14). For BM, which is continuous both in space and time, the probabilistic interpretation of $p_1(\rho, r)$ exactly at $\rho = 0$, $p_1(0, r)$, is a bit ill defined. Indeed, roughly speaking, $p_1(0, r)$ is the probability that the trajectory visits the points located at a distance within $[r, r + dr]$ from x_{max} only “a few times.” But we know that, if a site is visited once by BM, it will be visited again infinitely many times right after. As shown below, it is however possible to give a probabilistic interpretation to $p_1(0, r)$ by considering BM as a limit of a discrete lattice random walk (RW). We also

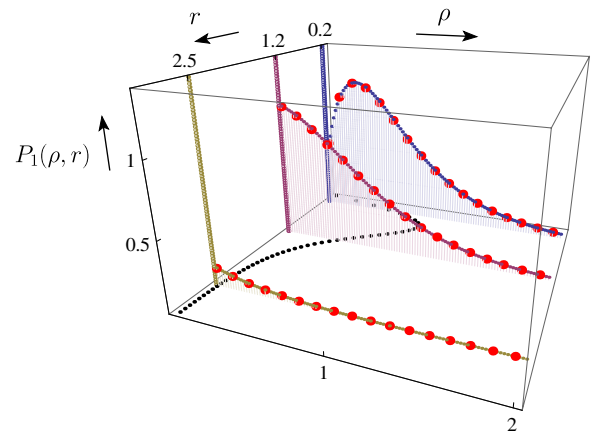


FIG. 2 (color online). Plot of $P_1(\rho, r)$ as a function of ρ for different values of r . The solid lines for $\rho = 0$ represent the $\delta(\rho)$ peak, $\propto \delta(\rho)$ in Eq. (12). The dotted lines correspond to our exact analytical results for the background density $p_1(\rho, r)$ in Eq. (12)—where the inverse LT with respect to s has been performed numerically (in green for $r = 2.5$, purple for $r = 1.2$, and blue for $r = 0.2$ as indicated on the r axis)—while the red dots indicate the results of simulations. On the $z = 0$ plane, we have plotted the exact mean DOS in Eq. (2).

obtain an exact expression for the moments of arbitrary order $k \in \mathbb{N}$, $\mu_k(r, t) = \langle [\rho(r, t)]^k \rangle$ given in Eq. (11). Finally, we show that our method can be extended to study the DOS of constrained BMs, like the Brownian bridge (BB), i.e., BM starting and ending at the origin.

Free BM.—To study analytically the PDF of $\rho(r, t)$, we compute its LT, $\langle e^{-\lambda \rho(r, t)} \rangle$. This is a particular functional of x_{\max} , of the form $\langle \exp[-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))] \rangle$. In our case (1) $V(y) = \delta(y - r)$ but the path integral method that we develop below holds actually for any arbitrary function $V(y)$. Denoting by t_{\max} the time at which the maximum is reached, the two intervals $[0, t_{\max}]$ and $[t_{\max}, t]$ are statistically independent (as BM is Markovian), and the PDF of t_{\max} is given by the arcsine law $P(t_{\max}) = 1/[\pi \sqrt{t_{\max}(t - t_{\max})}]$ [23]. The process $y(\tau) = x_{\max} - x(\tau)$ is obviously a BM which stays positive on $[0, t]$. By reversing the time arrow in the interval $[0, t_{\max}]$ and taking t_{\max} as the new origin of time, we see that $y(\tau)$ is built from two independent Brownian meanders (BMe)s: one of duration t_{\max} and the other (independent) one of duration $t - t_{\max}$ (see the Supplemental Material [27]). We recall that a BMe of duration T is a BM, starting at the origin, staying positive on the time interval $[0, T]$ and ending anywhere on the positive axis at time T . Therefore one has

$$\langle e^{-\lambda \int_0^t d\tau V[x_{\max} - x(\tau)]} \rangle = \int_0^t dt_{\max} \varphi(t_{\max}) \varphi(t - t_{\max}), \quad (5)$$

$$\varphi(\tau) = \frac{1}{\sqrt{\pi\tau}} \langle e^{-\lambda \int_0^\tau du V[y(u)]} \rangle_+, \quad (6)$$

where $\langle \cdots \rangle_+$ denotes an average over the trajectories of a BMe $y(\tau)$. In Eq. (6) the prefactor $1/\sqrt{\pi\tau}$ comes from the PDF of t_{\max} . This functional of the BMe $\varphi(\tau)$ can then be computed using path-integral techniques [28], which needs to be suitably adapted to our case. Indeed, for a BMe, which is continuous both in space and time, it is well known that one cannot impose simultaneously $y(0) = 0$ and $y(0^+) > 0$. This can be circumvented [29,30] by introducing a cutoff $\varepsilon > 0$ such that $y(0) = \varepsilon$ and then taking eventually the limit $\varepsilon \rightarrow 0$ of the following ratio defining $\varphi(\tau)$ in Eq. (6):

$$\langle e^{-\lambda \int_0^\tau du V[y(u)]} \rangle_+ = \lim_{\varepsilon \rightarrow 0} \frac{\int_0^\infty \langle y_F | e^{-H_\lambda \tau} | \varepsilon \rangle dy_F}{\int_0^\infty \langle y_F | e^{-H_0 \tau} | \varepsilon \rangle dy_F}, \quad (7)$$

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dx^2} + \lambda V(x) + V_{\text{wall}}(x), \quad (8)$$

where $V_{\text{wall}}(x)$ is a hard-wall potential, $V_{\text{wall}}(x) = 0$ for $x \geq 0$, and $V_{\text{wall}}(x) = +\infty$ for $x < 0$, which guarantees that the walker stays positive, as it should for a BMe. Note that in Eq. (7), y_F denotes the final point of the BMe, which can be anywhere on the positive axis. The convolution structure of the expression in Eq. (5) suggests to compute its LT with respect to t . Using the above result

in Eq. (7) applied to $V(x) = \delta(x - r)$ one finds, after some manipulations (see the Supplemental Material [27])

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \rho(r, t)} \rangle = \frac{1}{s} \left(\frac{\sqrt{2s} + \lambda(1 - e^{-\sqrt{2s}r})^2}{\sqrt{2s} + \lambda(1 - e^{-2\sqrt{2s}r})} \right)^2. \quad (9)$$

The expansion of Eq. (9) in powers of λ yields the LT of the moments $\tilde{\mu}_k(r, s) = \int_0^\infty \mu_k(r, t) e^{-st} dt$, for $k \in \mathbb{N}$. To invert these LTs, we introduce the family of functions $\Phi^{(j)}$, $j \in \mathbb{N}$, which satisfy

$$\frac{e^{-\sqrt{2s}u}}{(\sqrt{2s})^{j+1}} = \int_0^\infty t^{(j-1)/2} \Phi^{(j)}\left(\frac{u}{\sqrt{t}}\right) e^{-st} dt. \quad (10)$$

These functions can be obtained explicitly by induction, using $\Phi^{(0)}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $\Phi^{(j+1)}(x) = \int_x^\infty \Phi^{(j)}(u) du$ [31,32]. In terms of the $\Phi^{(j)}$'s (10), we obtain

$$\begin{aligned} \mu_k(r, 1) = 8k! \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} & [(2l+1)\Phi^{(k+1)}((2l+1)r) \\ & + (k-2(l+1))\Phi^{(k+1)}(2(l+1)r)]. \end{aligned} \quad (11)$$

For $k = 1$, this yields the result in Eq. (2), using $\Phi^{(2)}(x) = h(x)$ in Eq. (2). By inverting the LT with respect to λ in Eq. (9), we obtain

$$\begin{aligned} & \int_0^\infty e^{-st} P_t(\rho, r) dt \\ &= \delta(\rho) \frac{(e^{-\sqrt{2s}r} - 1)^2}{s(1 + e^{-\sqrt{2s}r})^2} + \frac{e^{-\rho\sqrt{2s}e^{\sqrt{2s}r}/(2\sinh(\sqrt{2s}r))}}{\cosh^3(\frac{r\sqrt{2s}}{2})} \\ & \times \left(\frac{e^{r\sqrt{2s}/2}}{\sqrt{2s}} + \frac{\rho e^{\sqrt{2s}r}}{4 \sinh(r\sqrt{2s}) \sinh(\frac{r\sqrt{2s}}{2})} \right), \end{aligned} \quad (12)$$

which has a much more complicated analytical structure than the corresponding LT of the PDF of the local time $T_{\text{loc}}(r, t)$ of free BM [33].

After Laplace inversion with respect to s of Eq. (12), one obtains the formula announced in Eq. (3). Indeed we can check that the coefficient of the term $\propto \delta(\rho)$ (12) is the LT with respect to t of

$$F_W(r, t) = 1 + \sum_{l=1}^\infty 4l(-1)^l \text{erfc}(lr/\sqrt{2t}), \quad (13)$$

which corresponds precisely to the distribution of the width of BM [34]. The second term, which is the LT with respect to t of $p_t(\rho, r)$ [35], has a more complicated structure. By analyzing it for small and large ρ we obtain the asymptotic behaviors given in Eq. (4). In particular the limiting function $p_t(0, r) = \lim_{\rho \rightarrow 0} p_t(\rho, r)$ in Eq. (4) is given by (see the Supplemental Material [27])

$$p_t(0, r) = \frac{1}{2} \delta(r) + 2 \sqrt{\frac{2}{\pi t}} \sum_{l=0}^{\infty} (-1)^{l+1} l(l+1) e^{-l^2 r^2 / (2t)}, \quad (14)$$

such that $\lim_{r \rightarrow 0^+} p_t(0, r) = \frac{1}{\sqrt{2\pi t}}$. As we explained it above, the meaning of $p_t(\rho = 0, r)$ (14) is a bit unclear for BM. One can however make sense of this quantity by considering BM as the scaling limit of a lattice RW of n steps, when $n \rightarrow \infty$. In particular, we can show that in this case the delta peak, $\propto \delta(r)$, in $p_t(0, r)$ (14) corresponds to trajectories with a unique maximum. The amplitude $1/2$ in front of this delta peak implies that, when $n \rightarrow \infty$, the probability that the RW has a unique maximum is $1/2$. This result can also be checked by an independent direct calculation. For such a lattice RW, it is also possible (see the Supplemental Material [27]) to give a probabilistic interpretation to the infinite sum in Eq. (14). Finally, in Fig. 2 we show the results of $p_1(\rho, r)$ obtained from numerical simulations (averages are performed over 10^7 samples) for three different values of r . We see that they are in perfect agreement with our exact formula (12).

Brownian bridge.—In the case of a BB, the method presented above can be straightforwardly adapted to compute the PDF of the DOS $\rho^B(r, t)$ with the simple modification that the PDF of t_{\max} is now uniform (a consequence of periodic boundary conditions). There is however a simpler way to do this calculation by mapping $\rho^B(r, t)$ to the (standard) local time of a Brownian excursion (BE), which is a BB conditioned to stay positive. To construct this mapping, we first transform the path by considering $y(\tau) = x_{\max} - x(\tau)$. We then break the time interval into two parts $[0, t_{\max}]$ and $[t_{\max}, t]$ and permute the two associated portions of the path, the continuity of the path being guaranteed by $x(t) = x(0) = 0$. We finally take the origin of times at t_{\max} to obtain a BE $x_E(\tau)$ on the interval $[0, t]$. This construction is well known in the literature under the name of Vervaat's transformation [38]. This shows that $\rho^B(r, t)$ is identical in law to the local time $T_{\text{loc}}^E(r, t)$ in r [23]

$$\rho^B(r, t) \stackrel{\text{law}}{=} T_{\text{loc}}^E(r, t) = \int_0^t \delta[x_E(\tau) - r] d\tau, \quad (15)$$

for the BE $x_E(\tau)$. By performing a similar transformation, substituting t_{\max} by t_{\min} —the time at which the minimum is reached—we can show that $\rho^B(r, t)$ for BB and $\rho^E(r, t)$ for BE are identical in law.

The LT of the PDF of $T_{\text{loc}}^E(r, t)$ in Eq. (15), $\langle e^{-\lambda T_{\text{loc}}^E(r, t)} \rangle_E$, where $\langle \cdots \rangle_E$ refers to the average over the BE, can be computed using path integral techniques. As explained above in Eq. (7) we introduce a cutoff $\varepsilon > 0$ such that $x_E(0) = x_E(t) = \varepsilon$ and obtain $\langle e^{-\lambda T_{\text{loc}}^E(r, t)} \rangle_E$ as

$$\langle e^{-\lambda T_{\text{loc}}^E(r, t)} \rangle_E = \lim_{\varepsilon \rightarrow 0} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle}, \quad (16)$$

where H_λ is given in Eq. (8) with $V(x) = \delta(x - r)$. The spectrum of H_λ can be computed and one obtains

$$\langle e^{-\lambda T_{\text{loc}}^E(r, t)} \rangle = \int_0^\infty \frac{dk \sqrt{\frac{2t}{\pi}} k^2 e^{-k^2 t/2}}{1 + \frac{4\lambda}{k} \sin(kr) \left(\frac{\lambda}{k} \sin(kr) + \cos(kr) \right)}. \quad (17)$$

By studying the large λ behavior of Eq. (17), which is of order $\mathcal{O}(\lambda^0)$, we can show that the PDF of $\rho^B(r, t)$ has an expression similar to, albeit different from, the one for BM in Eq. (3): $P_t^B(\rho, r) = F_W^B(r, t) \delta(\rho) + p_t^B(\rho, r)$, where $F_W^B(r, t)$ is the distribution function of the width of the BB [39] $F_W^B(r, t) = 1 + 2 \sum_{l=1}^\infty e^{-2l^2 r^2 / t} [1 - (4l^2 r^2 / t)]$, while $p_t^B(\rho, r)$ is now a different distribution.

Although the moments $\mu_k^B(r, t) = \langle [\rho^B(r, t)]^k \rangle$ can be obtained from Eq. (17), there is a much simpler way to compute them by using the mapping to $T_{\text{loc}}^E(r, t)$ of a BE (15). One has indeed $\mu_k^B(r, t) = \langle \prod_{i=1}^k \int_0^t dt_i \delta[x_E(t_i) - r] \rangle_E$, which can be written as convolutions of propagators of the BE. This calculation can be performed to get [see also Eq. (38) in Ref. [40]]

$$\mu_k^B(r, t) = 2\sqrt{2\pi k!} \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} \Phi^{(k-2)}[2r(l+1)], \quad (18)$$

with $\Phi^{(-1)} = -d\Phi^{(0)}/dr$ and where the $\Phi^{(j)}$'s have been defined below Eq. (10). For $k = 1$, one finds the mean DOS for the BB on the unit time interval, $\bar{\rho}^B(x) = \mu_1^B(x, 1) = 4xe^{-2x^2}$, as found in Ref. [40]. Note that it coincides in this case with the PDF of the maximum of a BB [which is a generic property for periodic signals such that $x(t) = x(0)$ [41]].

One can also show that Eq. (18) yields back the complicated though explicit formula for $p_t^B(\rho, r)$ found in Ref. [40,42] using a completely different method. In particular, for large ρ , one finds $p_1^B(\rho, r) \sim 16\rho^3 e^{-(\rho+2r)^2/2}$ [40,42], slightly different from Eq. (4) for BM, while $\lim_{\rho \rightarrow 0} p_1^B(\rho, r) = p_1^B(0, r)$ where $p_1^B(0, r) = \frac{1}{2} \delta(r) + \frac{1}{2} \partial_r F_W^B(r, t)$. This formula can be interpreted exactly as we did for BM [see below Eq. (14)].

Discussion.—The method developed here, in particular the formulas in Eqs. (5)–(7), is very general and can be used to study any functional of x_{\max} . Here we have studied the case of $\rho(r, t)$ in Eq. (1), which corresponds to $V(x) = \delta(x - r)$ but another class of functionals of x_{\max} , with several applications, are of the form $T_\alpha(t) = \int_0^t [x_{\max} - x(\tau)]^\alpha d\tau$. They correspond to a potential $V(x) = x^\alpha$ in Eqs. (7) and (8). The case $\alpha = -1$ is quite interesting as $T_{-1}(t)$ describes the fluctuations of the cost of the optimal search algorithm for the maximum of a RW [43,44]. The case $\alpha = -1/2$ is also interesting as $T_{-1/2}(t)$ describes the largest exit time of a particle diffusing through a random (Brownian) potential. Finally the case $\alpha = 1$ corresponds

to the area under a Brownian meander [29,30]. Our method (5)–(7) allows us to study the statistics of $T_\alpha(t)$ for any value of α , interpolating between the aforementioned observables, using a unifying physical approach. Besides the potential applications of the method developed here, our exact results for near-extreme statistics of a strongly correlated process as BM, gives rise to further challenging questions. The first one concerns the temporal resolution of the density of near extremes. While in Eq. (1) we have studied a time integrated observable, it is natural to study the statistics of the quantity $\delta[x_{\max} - x(\tau) - r]$ and its correlations at different values of τ . Finally it will be interesting to extend the present results to other stochastic processes, like for instance Lévy flights or branching processes.

We thank C. Banderier for useful correspondence and S. Sabhapandit for useful discussions at the earliest stage of this work. We also acknowledge support by ANR Grant No. 2011-BS04-013-01, WALKMAT, and in part by the Indo-French Centre for the Promotion of Advanced Research under Project No. 4604-3. G.S acknowledges partial support from the Labex PALM (project RANDMAT).

*anthony.perret@lptms.u-psud.fr

†comtet@lptms.u-psud.fr

‡majumdar@lptms.u-psud.fr

§gregory.schehr@lptms.u-psud.fr

- [1] E. J. Gumbel, *Statistics of Extremes* (Columbia University, New York, 1958).
- [2] R. W. Katz, M. P. Parlange, and P. Naveau, *Adv. Water Resour.* **25**, 1287 (2002).
- [3] P. Embrecht, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer-Verlag, Berlin, 1997).
- [4] S. N. Majumdar and J.-P. Bouchaud, *Quant. Finance* **8**, 753 (2008).
- [5] J.-P. Bouchaud and M. Mézard, *J. Phys. A* **30**, 7997 (1997).
- [6] D. S. Dean and S. N. Majumdar, *Phys. Rev. E* **64**, 046121 (2001).
- [7] P. Le Doussal and C. Monthus, *Physica (Amsterdam)* **317A**, 140 (2003).
- [8] D. S. Fisher and D. A. Huse, *Phys. Rev. B* **38**, 373 (1988).
- [9] D. S. Fisher and D. A. Huse, *Phys. Rev. B* **38**, 386 (1988).
- [10] C. Monthus and P. Le Doussal, *Eur. Phys. J. B* **41**, 535 (2004).
- [11] H. A. David and H. N. Nagaraja, *Order Statistics* (Wiley, New York, 2003), 3rd ed.
- [12] E. Brunet and B. Derrida, *Europhys. Lett.* **87**, 60010 (2009).
- [13] E. Brunet and B. Derrida, *J. Stat. Phys.* **143**, 420 (2011).
- [14] N. R. Moloney, K. Ozogány, and Z. Rácz, *Phys. Rev. E* **84**, 061101 (2011).
- [15] G. Schehr and S. N. Majumdar, *Phys. Rev. Lett.* **108**, 040601 (2012).
- [16] S. N. Majumdar, P. Mounaix, and G. Schehr, *Phys. Rev. Lett.* **111**, 070601 (2013).
- [17] F. Omori, *J. Coll. Sci., Imp. Univ. Tokyo* **7**, 111 (1894).
- [18] D. Vere-Jones, *Bull. Seismol. Soc. Am.* **59**, 1535 (1969).
- [19] A. M. Petersen, F. Wang, S. Havlin, and H. E. Stanley, *Phys. Rev. E* **82**, 036114 (2010).
- [20] S. Sabhapandit and S. N. Majumdar, *Phys. Rev. Lett.* **98**, 140201 (2007).
- [21] Ph. Ciais *et al.*, *Nature (London)* **437**, 529 (2005).
- [22] S. Sabhapandit, S. N. Majumdar, and S. Redner, *J. Stat. Mech.* (2008) L03001.
- [23] P. Lévy, *Compos. Math.* **7**, 283 (1940).
- [24] A. G. Pakes and F. W. Steutel, *Australian J. Stat.* **39**, 179 (1997).
- [25] A. G. Pakes and Y. Li, *Stat. Probab. Lett.* **40**, 395 (1998).
- [26] M. Politi, N. Millot, and A. Chakraborti, *Physica (Amsterdam)* **391A**, 147 (2012).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.111.240601> for details.
- [28] S. N. Majumdar, *Curr. Sci.* **89**, 2076 (2005).
- [29] S. N. Majumdar and A. Comtet, *Phys. Rev. Lett.* **92**, 225501 (2004).
- [30] S. N. Majumdar and A. Comtet, *J. Stat. Phys.* **119**, 777 (2005).
- [31] They can be written as $\Phi^{(j)}(x) = p_j(x) \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} + q_j(x) \operatorname{erfc}(\frac{x}{\sqrt{2}})$ where p_j and q_j are rational polynomials of degree $j-2$ and $j-1$, respectively, for $j \geq 2$ [32].
- [32] P. Chassaing and G. Louchard, *J. Algorithms* **44**, 29 (2002).
- [33] J. Pitman, *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1999), Vol. 1709, pp. 388–394.
- [34] W. Feller, *Ann. Math. Stat.* **22**, 427 (1951).
- [35] Note that this expression [Eq. (12)] is similar though different from the one found for the local time of a BE [36,37].
- [36] G. Hooghiemstra, *Proc. Am. Math. Soc.* **84**, 127 (1982).
- [37] B. Gittenberger and G. Louchard, *J. Appl. Probab.* **36**, 350 (1999).
- [38] W. Vervaat, *Ann. Probab.* **7**, 143 (1979).
- [39] K. L. Chung, *Ark. Mat.* **14**, 155 (1976).
- [40] L. Takács, *J. Appl. Math. Stoch. Anal.* **8**, 209 (1995).
- [41] T. W. Burkhardt, G. Györgyi, N. R. Moloney, and Z. Rácz, *Phys. Rev. E* **76**, 041119 (2007).
- [42] L. Takács, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5011 (1992).
- [43] A. M. Odlyzko, *Random Struct. Algorithms* **6**, 275 (1995).
- [44] P. Chassaing, J. F. Marckert, and M. Yor, *Ann. Appl. Probab.* **13**, 1264 (2003).

Near-extreme eigenvalues and the first gap of hermitian random matrices.

[126] A. Perret et G. Schehr (2014).
Journal of Statistical Physics, 156(5) :843–876.
Post-print version

Anthony Perret · Grégory Schehr

Near-extreme eigenvalues and the first gap of Hermitian random matrices

July 1, 2015

Abstract We study the phenomenon of “crowding” near the largest eigenvalue λ_{\max} of random $N \times N$ matrices belonging to the Gaussian Unitary Ensemble (GUE) of random matrix theory. We focus on two distinct quantities: (i) the density of states (DOS) near λ_{\max} , $\rho_{\text{DOS}}(r, N)$, which is the average density of eigenvalues located at a distance r from λ_{\max} and (ii) the probability density function of the gap between the first two largest eigenvalues, $p_{\text{GAP}}(r, N)$. In the edge scaling limit where $r = \mathcal{O}(N^{-1/6})$, which is described by a double scaling limit of a system of unconventional orthogonal polynomials, we show that $\rho_{\text{DOS}}(r, N)$ and $p_{\text{GAP}}(r, N)$ are characterized by scaling functions which can be expressed in terms of the solution of a Lax pair associated to the Painlevé XXXIV equation. This provides an alternative and simpler expression for the gap distribution, which was recently studied by Witte, Bornemann and Forrester in *Nonlinearity* **26**, 1799 (2013). Our expressions allow to obtain precise asymptotic behaviors of these scaling functions both for small and large arguments.

1 Introduction and summary of main results

Extreme value statistics (EVS) is currently an active subject of studies in various areas of sciences, and in particular in statistical physics [1–3]. The standard question which is usually addressed concerns the fluctuations of the maximum X_{\max} among a collections of N random variables X_1, \dots, X_N . However in many circumstances, it is natural to wonder about the “crowding” near the maximum X_{\max} : is X_{\max} very far from the others or, on the contrary, are there many others close to it [4] ? This type of question arises for instance naturally in the study of complex and disordered systems where the thermodynamical properties are dominated by the low-lying states, close to the ground state [5–8]. Such questions about near extreme events play also an important role in natural sciences [9, 10] or in finance [11]. They were also studied in the context of sporting events, like in marathon packs [12].

A natural way to characterize quantitatively this phenomenon of crowding is to study the full order statistics [13] where one considers not only the first maximum $M_{1,N} = X_{\max}$ but also the second $M_{2,N}$, the third $M_{3,N}$, \dots , more generally the k^{th} maximum $M_{k,N}$. A set of particularly interesting random variables which are sensitive to the crowding of the extremum are the gaps $d_{k,N} = M_{k,N} - M_{k+1,N}$ between successive maxima. While order (or gap) statistics have been widely studied in the past for independent and identically distributed (i.i.d.) random variables [13, 14], there exists very few exact results for strongly correlated random variables. Yet, many problems of statistical physics involve order

A. Perret
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France

G. Schehr
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France

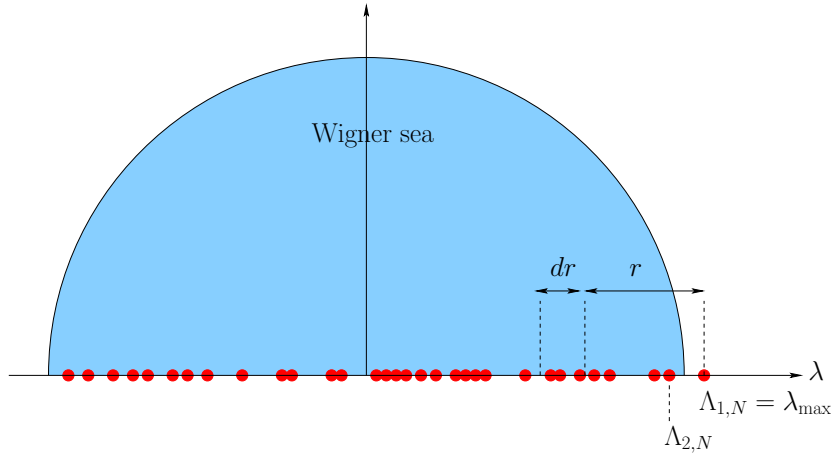


Fig. 1 Different quantities characterizing the “crowding” near the largest eigenvalue λ_{\max} studied in this paper: (i) the mean density of states $\rho_{\text{DOS}}(r, N)$ such that $\rho_{\text{DOS}}(r, N)dr$ is the mean number of eigenvalues located in the interval $[\lambda_{\max} - r - dr, \lambda_{\max} - r]$ and (ii) the PDF $p_{\text{GAP}}(r, N)$ of the spacing between the two largest eigenvalues, $p_{\text{GAP}}(r, N)dr = \text{Pr} .[(\Lambda_{1,N} - \Lambda_{2,N}) \in [r, r + dr]]$.

statistics of strongly correlated variables, as studied for branching Brownian motion [15, 16] or signals with $1/f^\alpha$ correlations [17]. Recent analytical progress was achieved for the gap statistics of random walks, which were shown to display a very rich structure [18, 19]. Another physically relevant instance of strongly correlated random variables where extreme value questions has attracted much attention is the set of eigenvalues of random matrices [25, 26]. In particular, the Tracy-Widom (TW) distributions [20, 21] which describe the fluctuations of the largest eigenvalues of random matrices belonging to the Gaussian ensembles (orthogonal, unitary and symplectic) of random matrix theory (RMT) are now cornerstones of the theory of EVS for strongly correlated random variables. For these ensembles, the distribution of the k^{th} eigenvalue can also be computed explicitly as rather simple generalization of the TW distribution [20–22] (we refer the reader to the Refs. [23, 24] for the analysis of these distributions in the large k limit where they become Gaussian). Despite the fact that one can formally write the joint probability density function (PDF) of the k first eigenvalues in terms of Fredholm determinants, it is only recently that the first gap between the first two largest eigenvalues of large Gaussian random matrices has been investigated in detail for GUE in Ref. [27] by Witte, Bornemann and Forrester (referred to as WBF in the following). In particular, they obtained an expression for the PDF of the first gap in terms of the components of a solution of a particular isomonodromic problem relating to the Painlevé II equation (see Appendix D). We will show below that the approach developed in this paper allows to obtain an alternative and somewhat simpler expression for this PDF of the first gap in the GUE ensemble.

Another natural quantity to characterize the phenomenon of crowding near the maximum is the (average) density of states (DOS) $\rho_{\text{DOS}}(r, N)$ near the maximum, which counts the mean number of random variables X_i ’s located at a given distance r from the maximum X_{\max} [4]:

$$\rho_{\text{DOS}}(r, N) = \frac{1}{N-1} \sum_{i \neq i_{\max}} \langle \delta(X_{\max} - X_i - r) \rangle, \quad (1)$$

where $X_{i_{\max}} = X_{\max}$ and $\langle \dots \rangle$ denotes an average over the different realizations of the random variables X_i ’s. This quantity $\rho_{\text{DOS}}(r, N)$ has been studied in mathematical statistics [29, 30] and, more recently, in physics [4], for i.i.d. and weakly correlated random variables. In particular, it was shown that, depending on whether the tail of the parent distribution of the X_i ’s decays slower than, faster than, or as a pure exponential, the limiting mean DOS converges to three different limiting forms. More recently, the DOS was computed exactly for Brownian motion [28] which is thus one rare instance of strongly correlated random variables where this phenomenon of crowding near extremes has been studied analytically.

Here we will focus on the set of eigenvalues of random matrices belonging to the Gaussian ensembles of RMT. Although the density of near extreme eigenvalues is a natural object to study, it has not – to our knowledge – been discussed before in the literature. Here we consider random Hermitian matrices, of size $N \times N$, belonging to the Gaussian Unitary Ensemble (GUE) of random matrices. The joint PDF of the eigenvalues is given by

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j} (\lambda_i - \lambda_j)^2 e^{-\sum_{i=1}^N \lambda_i^2}, \quad (2)$$

where the normalization constant is $Z_N = 2^{-\frac{N^2}{2}} (2\pi)^{\frac{N}{2}} \prod_{j=1}^N j!$. It is well known that the fluctuations of the eigenvalues are characterized by two different scales depending on their location in the spectrum: (i) in the bulk for $\lambda_i/\sqrt{N} = \mathcal{O}(1)$ and $|\lambda_i| < \sqrt{2N}$ and (ii) at the edge where $|\lambda_i \pm \sqrt{2N}| = \mathcal{O}(N^{-1/6})$. The existence of these two scales manifests itself in various observables associated to the eigenvalues of GUE (2), including their mean density defined as:

$$\rho(\lambda, N) = \frac{1}{N} \sum_{i=1}^N \langle \delta(\lambda_i - \lambda) \rangle. \quad (3)$$

One has obviously $\rho(\lambda, N) = \rho(-\lambda, N)$ and one can further show that, for large N , it exhibits two distinct regimes (for $\lambda > 0$) [25, 26, 31]

$$\rho(\lambda, N) \sim \begin{cases} \frac{1}{\sqrt{N}} \rho_{\text{bulk}}\left(\frac{\lambda}{\sqrt{N}}\right), & \lambda = \mathcal{O}(\sqrt{N}) \text{ \& } \lambda < \sqrt{2N}, \\ \sqrt{2N}^{-5/6} \rho_{\text{edge}}((\lambda - \sqrt{2N})\sqrt{2N}^{1/6}), & |\lambda - \sqrt{2N}| = \mathcal{O}(N^{-1/6}). \end{cases} \quad (4)$$

In Eq. (4), $\rho_{\text{bulk}}(x)$ is the Wigner semi-circle [25, 26]:

$$\rho_{\text{bulk}}(x) = \rho_W(x) = \frac{1}{\pi} \sqrt{2 - x^2}, \quad (5)$$

while $\rho_{\text{edge}}(x)$ is given by the Airy kernel at coinciding point [31] (see also Ref. [32]),

$$\rho_{\text{edge}}(x) = [\text{Ai}'(x)]^2 - x \text{Ai}^2(x), \quad (6)$$

whose asymptotic behaviors are given by ¹

$$\rho_{\text{edge}}(x) \sim \begin{cases} \frac{1}{\pi} \sqrt{-x}, & x \rightarrow -\infty, \\ \frac{1}{8\pi x} e^{-\frac{4}{3}x^{3/2}}, & x \rightarrow \infty. \end{cases} \quad (7)$$

Interestingly, one can check that these two regimes for $\rho(\lambda, N)$, the “bulk” one and the “edge” one in Eq. (4), perfectly match when λ approaches the value $\sqrt{2N}$ from below. Indeed, when $\lambda \rightarrow \sqrt{2N}$ from below, $\rho(\lambda, N)$ can be replaced by the Wigner semi-circle (5), which gives:

$$\rho(\lambda, N) \sim \frac{2^{3/4}}{\pi} N^{-3/4} \left(\sqrt{2N} - \lambda \right)^{1/2}, \quad \lambda \rightarrow \sqrt{2N}^-. \quad (8)$$

This behavior (8) coincides with the left tail of the scaling function $\rho_{\text{edge}}(x)$ in Eq. (7). Indeed, when the deviation from $\sqrt{2N}$ is large, $\sqrt{2N} - \lambda \sim \mathcal{O}(\sqrt{N})$, we can substitute in the second line of Eq. (4) the left tail asymptotic behavior of $\rho_{\text{edge}}(x)$ in (7), which gives

$$\rho(\lambda, N) \sim \sqrt{2N}^{-5/6} \frac{1}{\pi} \left(\sqrt{2N}^{1/6} (\sqrt{2N} - \lambda) \right)^{1/2}, \quad \lambda \rightarrow \sqrt{2N}^-, \quad (9)$$

¹ note that we have corrected a typo appearing in the large x asymptotic behavior of $\rho_{\text{edge}}(x)$ given in Eq. (3.11b) of Ref. [31].

which after a trivial rearrangement coincides with Eq. (8). Note also that the right tail of $\rho_{\text{edge}}(x)$ in Eq. (7) matches, as it should, with the right tail of the TW distribution for GUE [26] (see also [33]).

Here we focus on the density of eigenvalues near the largest one, $\lambda_{\text{max}} = \max_{1 \leq i \leq N} \lambda_i$ and consider, as in Eq. (1), the quantity $\rho_{\text{DOS}}(r, N)$ defined as (see Figs. 1 and 2)

$$\rho_{\text{DOS}}(r, N) = \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq i_{\text{max}}}}^N \langle \delta(\lambda_{\text{max}} - \lambda_i - r) \rangle, \quad (10)$$

where i_{max} is such that $\lambda_{i_{\text{max}}} = \lambda_{\text{max}}$ and $\langle \dots \rangle$ means an average taken with the weight in (2). It is normalized according to

$$\int_0^\infty dr \rho_{\text{DOS}}(r, N) = 1. \quad (11)$$

In this paper, we show how to compute this density $\rho_{\text{DOS}}(r, N)$, exactly for all N , in terms of an unconventional family of orthogonal polynomials (OP), which were introduced in Ref. [34] to provide a simple derivation of the Tracy-Widom distribution for GUE. Later on they were used in the context of multi critical matrix models in Ref. [36, 37]. They are unconventional in the sense that they are defined on a semi-infinite real interval (in the mathematical literature they are known to arise in the study of Janossy densities for unitary matrix ensembles [38, 39]). These OP are monic polynomials $\pi_k(\lambda, y)$ (which are polynomials of degree k of the variable λ , while y is a parameter) defined by [34]

$$\begin{cases} \langle \pi_k, \pi_{k'} \rangle = \int_{-\infty}^y d\lambda \pi_k(\lambda, y) \pi_{k'}(\lambda, y) e^{-\lambda^2} = \delta_{k,k'} h_k(y) \\ \pi_k(\lambda, y) = \lambda^k + \dots \end{cases} \quad (12)$$

By performing a large N analysis we show that for large N , $\rho_{\text{DOS}}(r, N)$ exhibits two different scaling regimes: (i) the bulk regime, where $r = \mathcal{O}(\sqrt{N})$ and (ii) the edge regime where $r = \mathcal{O}(N^{-1/6})$, which necessitates to study the aforementioned system of OP (12) in a double scaling limit which is made precise below. Our main results for $\rho_{\text{DOS}}(r, N)$ for large N can thus be summarized as follows

$$\rho_{\text{DOS}}(r, N) \sim \begin{cases} \frac{1}{\sqrt{N}} \tilde{\rho}_{\text{bulk}}\left(\frac{r}{\sqrt{N}}\right), & r = \mathcal{O}(\sqrt{N}) \text{ \& } 0 < r < 2\sqrt{2N}, \\ \sqrt{2}N^{-5/6} \tilde{\rho}_{\text{edge}}(r\sqrt{2}N^{1/6}), & r = \mathcal{O}(N^{-1/6}), \end{cases} \quad (13)$$

where $\tilde{\rho}_{\text{bulk}}(x)$ and $\tilde{\rho}_{\text{edge}}(\tilde{r})$ are scaling functions which we compute exactly. The first one is simply the Wigner semi-circle law with a shifted argument (see also Fig. 2)

$$\tilde{\rho}_{\text{bulk}}(x) = \rho_W(\sqrt{2} - x) = \frac{1}{\pi} \sqrt{x(2\sqrt{2} - x)}. \quad (14)$$

It can be simply understood from the definition of $\rho_{\text{DOS}}(r, N)$ in (10) as for large N , $\lambda_{\text{max}} \sim \sqrt{2N}$ while its fluctuations are of order $\mathcal{O}(N^{-1/6})$ and described by the TW distribution. Hence, if $r = \mathcal{O}(\sqrt{N})$ in (10) it is insensitive to the fluctuations of λ_{max} and its PDF can be simply replaced in (10) by a delta function $\delta(\lambda_{\text{max}} - \sqrt{2N})$. It thus follows that $\rho_{\text{DOS}}(r, N) \approx \rho(\sqrt{2N} - r, N)$ which using (5) yields the first line of (13) together with the expression in (14).

The expression of the scaling function $\tilde{\rho}_{\text{edge}}(x)$ has a more complicated and interesting analytical structure. We show that it can be written as

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx, \quad (15)$$

where $\mathcal{F}_2(u)$ is the TW distribution associated to GUE. We recall that it can be written as [20]

$$\mathcal{F}_2(x) = \exp \left[- \int_x^{\infty} (u - x) q^2(u) du \right], \quad (16)$$

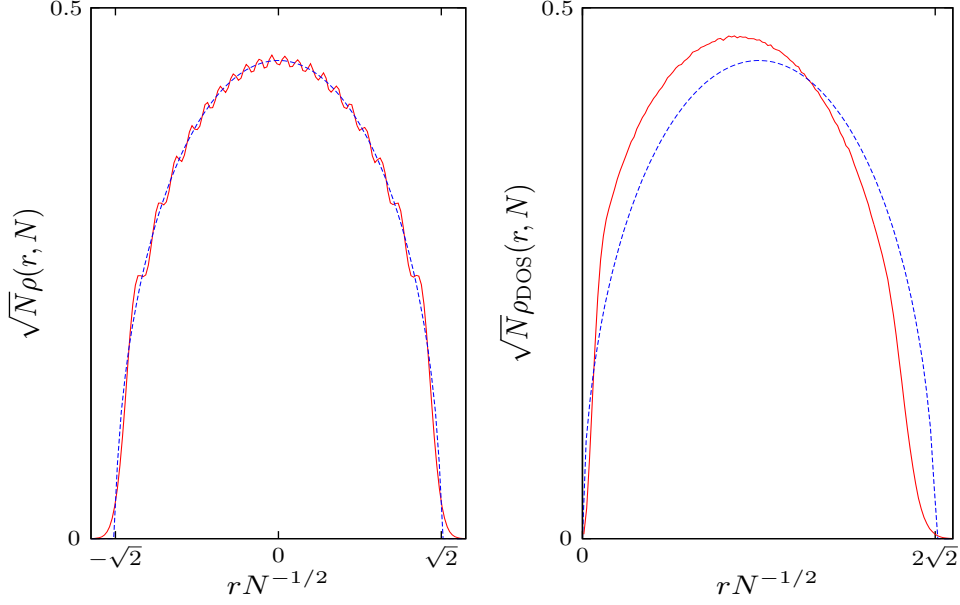


Fig. 2 Plots of the density of eigenvalues $\rho(r, N)$ and $\rho_{\text{DOS}}(r, N)$, evaluated numerically for GUE matrices of sizes $N = 20$ (red solid curve) and the two asymptotic bulk behavior (blue dashed curve).

where $q(x)$ in Eqs. (15) and (16) is the Hastings-McLeod solution of Painlevé II with parameter $\alpha = 0$ ²

$$\begin{cases} q'' = 2q^3 + xq, \\ q(x) \sim \text{Ai}(x) \text{ for } x \rightarrow \infty, \end{cases} \quad (17)$$

with $q \equiv q(x)$ and where $\text{Ai}(x)$ is the Airy function. In Eq. (15) the function $\tilde{f}(\tilde{r}, x)$ satisfies a Schrödinger like equation with a prescribed asymptotic behavior

$$\partial_x^2 \tilde{f}(\tilde{r}, x) - [x + 2q^2(x)] \tilde{f}(\tilde{r}, x) = -\tilde{r} \tilde{f}(\tilde{r}, x), \quad \tilde{f}(\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} 2^{-1/6} \sqrt{\pi} \text{Ai}(x - \tilde{r}). \quad (18)$$

We show below that $\tilde{f}(\tilde{r}, x)$ is the first component of the $2d$ -vector (namely a psi-function) $(\tilde{f}(\tilde{r}, x), \tilde{g}(\tilde{r}, x))$ which is a solution of the Lax pair associated to the Painlevé XXXIV equation [43] (with parameter $\alpha = 0$). Indeed $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ are solutions of the system of differential equations

$$\frac{\partial}{\partial \tilde{r}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{A}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad \frac{\partial}{\partial x} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{B}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad (19)$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are 2×2 matrices given by

$$\tilde{\mathbf{A}} = \begin{pmatrix} -\frac{q'(x)}{q(x)} & 1 + \frac{q^2(x)}{\tilde{r}} \\ -\tilde{r} - \frac{R(x)}{q^2(x)} & \frac{q'(x)}{q(x)} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \frac{q'(x)}{q(x)} & -1 \\ \tilde{r} & -\frac{q'(x)}{q(x)} \end{pmatrix}, \quad (20)$$

with $R(x) = \int_x^\infty q^2(u) du$ and where the solutions $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ are characterized by the asymptotic behaviors [43]

$$\tilde{f}(\tilde{r}, x) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-1/6} \tilde{r}^{-1/4} \sin \left(\frac{2}{3} \tilde{r}^{3/2} - x \sqrt{\tilde{r}} + \frac{\pi}{4} \right) + \mathcal{O}(\tilde{r}^{-3/4}), \quad (21a)$$

$$\tilde{g}(\tilde{r}, x) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-1/6} \tilde{r}^{1/4} \cos \left(\frac{2}{3} \tilde{r}^{3/2} - x \sqrt{\tilde{r}} + \frac{\pi}{4} \right) + \mathcal{O}(\tilde{r}^{-1/4}). \quad (21b)$$

² The Painlevé II equation with parameter α reads $q''_\alpha = 2q_\alpha^3 + xq_\alpha - \alpha$, see also below in Eq. (62). The Hastings-McLeod solution has the asymptotic behavior $q_\alpha(x) \sim \alpha/x$, for $x \rightarrow \infty$.

From these expressions, we obtain the asymptotic behaviors of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in Eq. (15) as

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \begin{cases} \frac{1}{2}\tilde{r}^2 + a_4\tilde{r}^4 + \mathcal{O}(\tilde{r}^6) & , \tilde{r} \rightarrow 0 \\ \frac{\sqrt{\tilde{r}}}{\pi} & , \tilde{r} \rightarrow \infty , \end{cases} \quad (22)$$

where $a_4 = -0.196788\dots$ is given explicitly below (105). Using the large \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ (22), one can show, as we did before for the density of the eigenvalues $\rho(\lambda, N)$, that the edge and the bulk regimes of $\rho_{\text{DOS}}(r, N)$ perfectly match.

Interestingly, we also show that the mean DOS, $\rho_{\text{DOS}}(r, N)$, is related to the PDF of the gap between the two largest eigenvalues. Let us denote by $\lambda_{\text{max}} = \Lambda_{1,N} \geq \Lambda_{2,N} \geq \dots \geq \Lambda_{N,N}$ and by $d_{1,N} = \Lambda_{1,N} - \Lambda_{2,N}$ the first gap. Its PDF is denoted by $p_{\text{GAP}}(r, N)$, such that $\text{Pr}[d_{1,N} \in [r, r + dr]] = p_{\text{GAP}}(r, N)dr$. Although the DOS $\rho_{\text{DOS}}(N, r)$ in Eq. (10) is naturally defined for $r > 0$, our exact formula for finite N suggests a natural analytic continuation to the real negative axis of this function. Furthermore, we show the following identity:

$$p_{\text{GAP}}(r, N) = (N - 1)\rho_{\text{DOS}}(-r, N) . \quad (23)$$

Here we focus on the *typical* fluctuations of the gap, i.e. when $r = \mathcal{O}(N^{-1/6})$ ³, and show similarly to Eq. (13) that for large N ,

$$p_{\text{GAP}}(r, N) = \sqrt{2}N^{1/6}\tilde{p}_{\text{typ}}\left(r\sqrt{2}N^{1/6}\right) , \quad (24)$$

where $\tilde{p}_{\text{typ}}(\tilde{r})$ is given by

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(-\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx . \quad (25)$$

This formula (25) should be compared to the relatively more complicated formula obtained by WBF in Ref. [27] (given in Appendix D). In addition to show a direct connection with the Lax system associated to Painlevé XXXIV, our formula (25) is amenable to a rather precise asymptotic analysis. We show indeed the following asymptotic behaviors:

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \begin{cases} \frac{1}{2}\tilde{r}^2 + a_4\tilde{r}^4 + \mathcal{O}(\tilde{r}^6) , & r \rightarrow 0 \\ A \exp\left(-\frac{4}{3}\tilde{r}^{3/2} + \frac{8}{3}\sqrt{2}\tilde{r}^{3/4}\right) \tilde{r}^{-21/32} \left(1 - \frac{1405\sqrt{2}}{1536}\tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2})\right) , & r \rightarrow +\infty , \end{cases} \quad (26)$$

with the amplitude $A = 2^{-91/48}e^{\zeta'(-1)}/\sqrt{\pi}$, where $\zeta'(x)$ is the derivative of the Riemann zeta function, while the amplitude a_4 is given in (105). These behaviors should be compared to the distribution of the spacing in the *bulk* of the spectrum, given by the Gaudin-Mehta distribution and well approximated by the Wigner surmise [25, 26], $P_{\beta}(\tilde{r}) = A_{\beta}\tilde{r}^{\beta}e^{-B_{\beta}\tilde{r}^2}$ (with some constants A_{β} and B_{β}). In particular, the latter has a Gaussian tail, at variance with the stretched exponential behavior $\tilde{p}_{\text{typ}}(\tilde{r}) \sim e^{-\frac{4}{3}\tilde{r}^{3/2}}$ found here at the *edge* (26).

Note that the occurrence of the Painlevé XXXIV equation was previously noted in the study of the OP in Eq. (12). In particular, using Riemann-Hilbert techniques, it was shown in Ref. [38] that the Lax pair studied here enters naturally the expression of the associated Christoffel-Darboux kernel associated to the OP in (12) in the double scaling limit (see below). More recently, using the Lax method, it was shown in Ref. [36] that the Painlevé XXXIV hierarchy appears in the present OP system and its generalization to multi-critical matrix models. In this perspective, our main results are thus to show that (i) near extreme statistics, including order statistics, of GUE can naturally be expressed in terms of the OP given in Eq. (12) and (ii) provide a rather simple study of this OP system (12) in the double scaling limit, along the lines developed in theoretical physics [34, 40–42], where we will see that the Lax pair associated to Painlevé XXXIV emerges rather naturally.

³ *Atypically* large fluctuations would correspond to the case where $r = \mathcal{O}(\sqrt{N})$, which is not studied here.

The paper is organized as follows: in section 2, we give an explicit formula for $\rho_{\text{DOS}}(r, N)$ for any finite N in terms of the OP system in (12). In section 3 we perform a large N analysis of $\rho_{\text{DOS}}(N, r)$, both in the bulk and in the edge scaling limit – the latter corresponding to a double scaling limit of the OP system. In section 4, we focus on the typical fluctuations of the first gap, where $p_{\text{GAP}}(r, N)$ is analyzed in the limit when N is large and $r \sim \mathcal{O}(N^{-1/6})$. In section 5, we present a numerical evaluation of our formulas, providing in particular a direct comparison with the result of WBF [27], before we conclude in section 5. Some technical details have been left in Appendices A, B and C while the formula of WBF [27] for the PDF of the first gap has been given, for completeness, in Appendix D.

2 An exact formula for $\rho_{\text{DOS}}(r, N)$ for finite N

We start by deriving an exact formula for the density of states $\rho_{\text{DOS}}(r, N)$ valid for any finite N , in terms of the OP in (12). The obtained formula turns out to be useful for a large N asymptotic analysis.

2.1 Introducing orthogonal polynomials

First, we notice that, using the invariance of the joint PDF $P_{\text{joint}}(\lambda_1, \dots, \lambda_N)$ in (2) under any permutation of the λ_i 's, the mean DOS $\rho_{\text{DOS}}(r, N)$ can be written as

$$\rho_{\text{DOS}}(r, N) = N \int_{-\infty}^{\infty} dy \int_{-\infty}^y d\lambda_1 \int_{-\infty}^y d\lambda_2 \dots \int_{-\infty}^y d\lambda_{N-2} P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_{N-2}, y - r, y), \quad (27)$$

where y denotes the actual value of λ_{max} . To perform this multiple integral in (27) it is convenient to introduce the monic orthogonal polynomials which were briefly presented in the introduction (12) $\pi_k(\lambda, y)$ (which are polynomials of degree k of the variable λ , while y is a parameter) defined by [34]

$$\begin{cases} \langle \pi_k, \pi_{k'} \rangle = \int_{-\infty}^y d\lambda \pi_k(\lambda, y) \pi_{k'}(\lambda, y) e^{-\lambda^2} = \delta_{k,k'} h_k(y) \\ \pi_k(\lambda, y) = \lambda^k + \dots \end{cases} \quad (28)$$

When $y \rightarrow \infty$, these orthogonal polynomials reduce to the Hermite polynomials. For finite y , there is no closed formula for these orthogonal polynomials. The first polynomials $\pi_k(\lambda, y)$ can however be computed from (28), to obtain [34]:

$$\pi_0(\lambda, y) = 1 \quad (29)$$

$$\pi_1(\lambda, y) = \lambda + a$$

$$\pi_2(\lambda, y) = \lambda^2 + \lambda \left(\frac{a + \lambda}{1 - 2a(a + \lambda)} - \lambda \right) - 1 + \frac{1}{2 - 4a(a + \lambda)}$$

$$\begin{aligned} \pi_3(\lambda, y) = & \lambda^3 + [\lambda^2 2a(8a^2 - 4a^2 y^2 + 8ay - 8ay^3 - 3 - 4y^4) \\ & + 2\lambda(-12a^3 y - 10a^2 - 22a^2 y^2 - 9ay - 10ay^3 + 3) \\ & + a(-16a^2 + 4a^2 y^2 - 20ay + 8ay^3 + 5 - 4y^2 + 4y^4)] / (8a^3 y + 12a^2 + 16a^2 y^2 + 12ay + 8ay^3 - 4) \end{aligned}$$

where $a \equiv a(y)$ is given by

$$a = \frac{e^{-y^2}}{\sqrt{\pi}(1 + \text{erf}(y))}, \quad (30)$$

where $\text{erf}(y)$ denotes the error function. From (29) one also obtains the first norms

$$h_0 = e^{-y^2} \frac{1}{2a} \quad (31)$$

$$h_1 = e^{-y^2} \frac{1}{4} \left(\frac{1}{a} - 2a - 2y \right)$$

$$h_2 = e^{-y^2} \frac{2a^3 y + 3a^2 + 4a^2 y^2 + 3ay + 2ay^3 - 1}{4a(2a^2 + 2ay - 1)}$$

$$h_3 = e^{-y^2} \frac{-32a^4 + 4a^4 y^2 - 60a^3 y + 16a^3 y^3 + 29a^2 - 20a^2 y^2 + 20a^2 y^4 + 30ay + 8ay^3 + 8ay^5 - 6}{16a(2a^3 y + 3a^2 + 4a^2 y^2 + 3ay + 2ay^3 - 1)}$$

As one sees in Eq. (29), the expression of $\pi_k(\lambda, y)$ becomes more and more complicated as k grows and to analyze these polynomials for larger values of k , it is much more convenient to consider the three term recurrence relation which they satisfy [35]

$$\lambda \pi_k(\lambda, y) = \pi_{k+1}(\lambda, y) + S_k(y)\pi_k(\lambda, y) + R_k(y)\pi_{k-1}(\lambda, y), \quad (32)$$

$$\text{with } R_k(y) = \frac{h_k(y)}{h_{k-1}(y)}, \quad (33)$$

and $S_k(y) \neq 0$ as the interval of integration in (28) is not symmetric. The density $\rho_{\text{DOS}}(r, N)$ in (27) can then be expressed in terms of these orthogonal polynomials $\pi_k(\lambda, y)$. This is done by (i) replacing the Vandermonde determinant in P_{joint} (2, 27) by a determinant built from the OP $\pi_k(\lambda, y)$ and (ii) writing each determinant in terms of its Laplace expansion and use the orthogonality condition (28) to perform the integrals over $\lambda_1, \dots, \lambda_{N-2}$ in (27). By performing these standard manipulations [25, 26], we find that $\rho_{\text{DOS}}(r, N)$ can be written as

$$\rho_{\text{DOS}}(r, N) = \frac{N(N-2)!}{Z_N} \int_{-\infty}^{\infty} dy \prod_{k=0}^{N-1} h_k(y) \left| \begin{array}{cc} K_N(y-r, y-r) & K_N(y-r, y) \\ K_N(y, y-r) & K_N(y, y) \end{array} \right|, \quad (34)$$

where $K_N(\lambda, \lambda')$ is the kernel associated to the OP in (28), given by

$$K_N(\lambda, \lambda') = \sum_{k=0}^{N-1} \frac{1}{h_k(y)} \pi_k(\lambda, y) \pi_k(\lambda', y) e^{-\frac{\lambda^2 + \lambda'^2}{2}} = \sum_{k=0}^{N-1} \psi_k(\lambda, y) \psi_k(\lambda', y), \quad (35)$$

$$\text{where } \psi_k(\lambda, y) = \frac{1}{\sqrt{h_k(y)}} \pi_k(\lambda, y) e^{-\frac{\lambda^2}{2}}, \quad (36)$$

where we have introduced the normalized wave functions $\psi_k(\lambda, y)$. Using the Cristoffel-Darboux formula, the kernel $K_N(\lambda, \lambda')$ in (35) can be written as

$$K_N(\lambda, \lambda') = \sqrt{R_N(y)} \frac{\psi_N(\lambda, y) \psi_{N-1}(\lambda', y) - \psi_{N-1}(\lambda, y) \psi_N(\lambda', y)}{\lambda - \lambda'}, \quad (37)$$

which, of course, also depends on y . We note in passing that, although $\rho_{\text{DOS}}(r, N)$ is naturally defined for $r > 0$ [see Eq. (3)], the above expression (34) admits a natural extension for $r < 0$. This formula (34) can be explicitly evaluated for small values of N using the expressions in Eqs. (29) and (31). For instance, for $N = 4$, using these explicit expressions one obtains the expression of $\rho_{\text{DOS}}(r, 4)$ as a single integral which can easily be computed numerically. Eventually, as a check of our computation, this exact analytical expression can be compared to a numerical evaluation of this quantity, obtained by sampling a large number of 4×4 matrices. The result of this comparison is shown in Fig. 3 which shows a very nice agreement between our exact formula (29, 31, 34) – the black line on the figure – and the numerics – the blue points on that figure. We also notice that, compared to the density of eigenvalues $\rho(\lambda, N)$ (3), the DOS $\rho_{\text{DOS}}(r, N)$, for finite N , shows much less pronounced oscillations. This is also quite visible on Fig. 2 for larger values of N .

To conclude this paragraph, we emphasize that the exact expressions in (34) together with (37), which are valid for any value of N , constitute the starting point of the asymptotic large N analysis which we will perform in section 3. Before that, we derive a useful identity.

2.2 A useful identity and the normalization of $\rho_{\text{DOS}}(r, N)$

At this stage, it is useful to notice the following identity, whose derivation starts with the definition of the cumulative distribution of the largest eigenvalue $F_N(y) = \text{Pr.}(\max_{1 \leq i \leq N} \lambda_i \leq y)$:

$$F_N(y) = \int_{-\infty}^y d\lambda_1 \cdots \int_{-\infty}^y d\lambda_N P_{\text{joint}}(\lambda_1, \dots, \lambda_N). \quad (38)$$

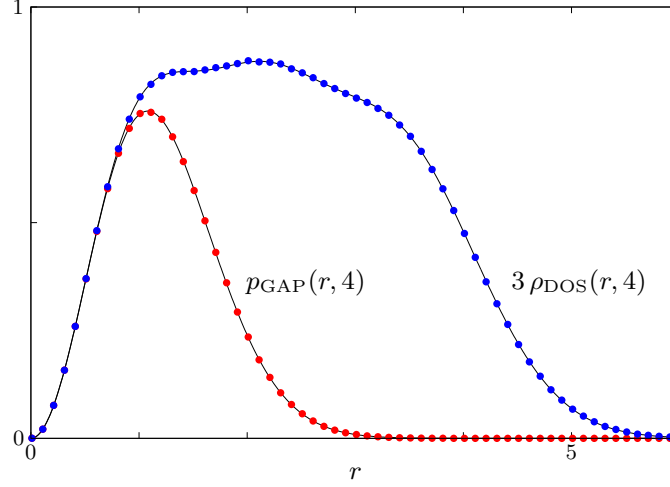


Fig. 3 Numerical result for a 4×4 GUE matrix (dots) compared to our analytical results – see Eqs. (23) and (34) – (black curve).

It can be evaluated by replacing the Vandermonde determinant by the determinant built from the OP in (28) and then use the Cauchy-Binet formula to obtain [25, 26, 34]

$$F_N(y) = \frac{N!}{Z_N} \prod_{j=0}^{N-1} h_j(y). \quad (39)$$

Differentiating Eq. (39) with respect to y , one obtains

$$\partial_y F_N(y) = \frac{N!}{Z_N} \sum_{k=0}^{N-1} \left(\partial_y h_k(y) \prod_{j=0, j \neq k}^{N-1} h_j \right) = \frac{N!}{Z_N} \left(\prod_{j=0}^{N-1} h_j(y) \right) \sum_{k=0}^{N-1} \frac{\partial_y h_k(y)}{h_k(y)}. \quad (40)$$

On the other hand, one has from (28)

$$\partial_y h_k(y) = [\pi_k(y, y)]^2 e^{-y^2} + 2 \langle \partial_y \pi_k, \pi_k \rangle = [\pi_k(y, y)]^2 e^{-y^2}, \quad (41)$$

as $\partial_y \pi_k(\lambda, y)$ is a polynomial in λ of degree $k-1$ [because $\pi_k(\lambda, y)$ is a monic polynomial of the variable λ (28)], which implies $\langle \partial_y \pi_k, \pi_k \rangle = 0$. Hence

$$\partial_y \ln F_N(y) = \sum_{k=0}^{N-1} \frac{[\pi_k(y, y)]^2}{h_k(y)} e^{-y^2} = K_N(y, y). \quad (42)$$

Using (39) and (42), one can write $\rho_{\text{DOS}}(r, N)$ in (34) as

$$\rho_{\text{DOS}}(r, N) = \frac{1}{N-1} \int_{-\infty}^{\infty} dy [F'_N(y) K_N(y-r, y-r) - F_N(y) K_N^2(y, y-r)], \quad (43)$$

where we have used that the kernel is symmetric $K_N(\lambda, \lambda') = K_N(\lambda', \lambda)$. Note that the second term in (43) is a contribution coming from the correlations between the largest eigenvalue λ_{max} in y and the eigenvalues at $y-r$. One can finally check the normalization of $\rho_{\text{DOS}}(r, N)$ by using the identities

$$\int_0^{+\infty} K_N(y-r, y-r) dr = \int_{-\infty}^y K_N(x, x) dx = N, \quad (44)$$

$$\int_0^{+\infty} K_N^2(y, y-r) dr = K_N(y, y) = \partial_y \ln F_N(y), \quad (45)$$

where we have used the identity (42) in (45). Note that from (44, 45) we see that the weight of the first term in (43) is of order $\mathcal{O}(1)$ (for large N) while the second term in (43) is of order $\mathcal{O}(N^{-1})$. We now proceed to the large N analysis of $\rho_{\text{DOS}}(r, N)$.

3 Large N asymptotics

As we have already pointed it out in the introduction [see for instance Eq. (4)], there are two different regimes characterizing the fluctuations of the eigenvalues: one is the *bulk* regime, for eigenvalues λ_i 's such that $\lambda_i/\sqrt{N} = \mathcal{O}(1)$ and $\lambda_i < \sqrt{2N}$, and the other scaling regime is at the *edge* where the eigenvalues are close to the edge of the Wigner semi-circle, $\lambda - \sqrt{2N} = \mathcal{O}(N^{-1/6})$. Hence when analyzing the large N behavior of $\rho_{\text{DOS}}(r, N)$, one expects two different scaling regimes: (i) when $r = \mathcal{O}(\sqrt{N})$ and (ii) when $r = \mathcal{O}(N^{-1/6})$. We now study these two regimes separately.

3.1 Regime (i): in the bulk when $r = \mathcal{O}(\sqrt{N})$

In this regime, it is easy to see that the second term in (43) is subdominant compared to the first one. We have already seen before that the total weight of the latter is only of order $\mathcal{O}(N^{-1})$ while the first term has a total weight of order $\mathcal{O}(1)$ [see our comment below Eqs. (44, 45)]. Physically this is also clear as the second one involves the correlations between λ_{max} , located in y , and the eigenvalues located in $y - r$, where in this regime (i), $\lambda_{\text{max}} - r = \mathcal{O}(\sqrt{N})$. To analyze the first term in (43) one can replace the PDF of λ_{max} , $\text{Pr.}(\lambda_{\text{max}} = y) = F'_N(y)$ by $\delta(y - \sqrt{2N})$: this is justified by the fact that $K_N(y - r, y - r)$ with $y - r = \mathcal{O}(\sqrt{N})$ is not sensitive to the typical fluctuations of λ_{max} around $y = \sqrt{2N}$ which are much smaller, of order $\mathcal{O}(N^{-1/6})$. Denoting $r = \hat{r}\sqrt{N}$, one thus has

$$\rho_{\text{DOS}}(\hat{r}\sqrt{N}, N) \sim \frac{1}{N} K_N(\sqrt{2N} - \hat{r}\sqrt{N}, \sqrt{2N} - \hat{r}\sqrt{N}) \underset{N \rightarrow \infty}{\sim} \frac{1}{\sqrt{N}} \tilde{\rho}_{\text{bulk}}\left(\hat{r} = \frac{r}{\sqrt{N}}\right),$$

$$\tilde{\rho}_{\text{bulk}}(\hat{r}) = \rho_W(\sqrt{2} - \hat{r}) = \frac{1}{\pi} \sqrt{\hat{r}(2\sqrt{2} - \hat{r})}, \quad (46)$$

which yields the result announced in the first line of (13) and in Eq. (14). As explained in the introduction, this result (46) can be easily understood as the density of eigenvalues at a distance r from $\lambda_{\text{max}} \sim \sqrt{2N}$ in (10) such that $r - \lambda_{\text{max}} = \mathcal{O}(\sqrt{N})$ is insensitive to the fluctuations of λ_{max} which are of order $\mathcal{O}(N^{-1/6})$. Hence, $\rho_{\text{DOS}}(r, N)$ is simply a shifted Wigner semi-circle (46). Note that this argument holds actually for any Gaussian β -ensemble where the joint PDF of the eigenvalues λ_i 's is given by Eq. (2) with the replacement of $\prod_{i,j} |\lambda_i - \lambda_j|^2$ by $\prod_{i,j} |\lambda_i - \lambda_j|^\beta$, where $\beta > 0$ can take any real value [25, 26]. In this case, one thus expects that the DOS behaves also as in Eq. (46). We now turn to the case where r is of order $\mathcal{O}(N^{-1/6})$.

3.2 Regime (ii): near the edge when $r = \mathcal{O}(N^{-1/6})$

In this regime, the analysis is more subtle. It requires the analysis of the kernel $K_N(y - r, y - r')$ when $y \sim \sqrt{2N}$ and both r, r' are of order $N^{-1/6}$. This corresponds to a double scaling limit analysis of the recursion relations for the OP in Eqs. (32) and (33). In Ref. [34], Nadal and Majumdar used this OP system (28) to compute the cumulative distribution function (CDF) of λ_{max} in Eq. (39), which can be expressed in terms of the norms $h_k(y)$'s only. To derive the Tracy-Widom distribution from (39), one needs indeed to perform the analysis of $h_k(y)$ for large k of order $\mathcal{O}(N)$, with $N \rightarrow \infty$ and, simultaneously, $y - \sqrt{2N} = \mathcal{O}(N^{-1/6})$. This double scaling analysis of the norms $h_k(y)$ was performed in Ref. [34] and generalized to the multi critical matrix models in Ref. [36]. Here, in addition to the norms, the computation of $\rho_{\text{DOS}}(r)$ in (34) requires the analysis of the kernel $K_N(y - x, y - x')$ in this double scaling limit when $N \rightarrow \infty$, with $y - \sqrt{2N} = \mathcal{O}(N^{-1/6})$ and with x and x' both of order $\mathcal{O}(N^{-1/6})$. To this purpose, we analyze and solve the three term recurrence relation in (32) in this double scaling limit.

3.2.1 Double scaling analysis of the three terms recurrence relation

We start with the recursion relation satisfied by the wave functions $\psi_N(\lambda, y)$ in Eq. (36), which is easily obtained from the recursion relation for the OP given in Eq. (32). It reads

$$\lambda \psi_N(\lambda, y) = \sqrt{R_{N+1}(y)} \psi_{N+1}(\lambda, y) + S_N(y) \psi_N(\lambda, y) + \sqrt{R_N(y)} \psi_{N-1}(\lambda, y), \quad (47)$$

where we recall that $\psi_N(\lambda, y) = \pi_N(\lambda, y)/\sqrt{h_N}e^{-\frac{\lambda^2}{2}}$, $R_N = h_N/h_{N-1}$ and $S_N = -\partial_y h_N(y)/2$. From Ref. [34], we know that, in the double scaling limit when $N \rightarrow \infty$ and $y - \sqrt{2N} = \mathcal{O}(N^{-1/6})$ the coefficients $R_N(y)$ and $S_N(y)$ take the scaling form

$$R_N(y) = \frac{N}{2} \left(1 - N^{-\frac{2}{3}} q^2(x) + \mathcal{O}(N^{-1}) \right), \quad S_N(y) = -\frac{N^{-\frac{1}{6}}}{\sqrt{2}} q^2(x) + \mathcal{O}(N^{-\frac{1}{2}}), \quad (48)$$

$$y = \sqrt{2N} + \frac{x}{\sqrt{2}} N^{-\frac{1}{6}},$$

where $q(x)$ is the Hastings-McLeod solution of the Painlevé II equation (17). Given the scaling form for $R_N(y)$ and $S_N(y)$ in (48), we search for a solution of the recursion relations in (47) under the scaling form

$$\psi_N(y - r, y) = \frac{2^{1/4}}{\sqrt{\pi}} N^\nu G \left(\sqrt{2} N^{1/6} r, \sqrt{2} N^{1/6} (y - \sqrt{2N}) \right), \quad (49)$$

where ν and the function G are still to be determined and where the amplitude $2^{1/4}/\sqrt{\pi}$ is chosen here for convenience. As mentioned above, the choice of the dependence of $\psi_N(y - r, y)$ in the scaling variable $\sqrt{2} N^{1/6} (y - \sqrt{2N})$ follows naturally from the dependence of the $R_N(y)$ and $S_N(y)$ in (47), (48) [34]. The dependence on the scaling variable $\sqrt{2} N^{1/6} r$ is a priori less obvious on the recurrence relation itself (47). It is however motivated by the analysis of the OP system (28) in the limit where $y \gg \sqrt{2N}$. Indeed, for $y \gg \sqrt{2N}$, one can replace y by $+\infty$ in the integral defining the OP $\pi_k(\lambda, y)$ in (28) such that the OP can be expressed in terms of the Hermite polynomials. This yields, for $y \gg \sqrt{2N}$ [34]:

$$\psi_N(\lambda, y) = \frac{1}{\pi^{1/4} 2^{N/2} \sqrt{N!}} H_N(\lambda) e^{-\frac{\lambda^2}{2}} + \mathcal{O}(e^{-y^2}), \quad (50)$$

where $H_N(\lambda)$ is the Hermite polynomial of degree N . We now want to evaluate the wave function $\psi_N(y - r, y)$ when y reaches the value $\sqrt{2N}$ from above, $y \rightarrow \sqrt{2N}^+$. Setting $y = \sqrt{2N} + x N^{-1/6}/\sqrt{2}$ as in (48) one finds, from (50), using the Plancherel-Rotach formula for Hermite polynomials [35] that $\psi_N(y - r, y)$ approaches the limiting form

$$\psi_N(y - r, y) \underset{N \rightarrow \infty}{\sim} 2^{1/4} N^{-1/12} \text{Ai}(x - \sqrt{2} N^{1/6} r) + \mathcal{O}(e^{-y^2}), \quad (51)$$

which is thus a function of the two scaling variables x in (48) and

$$\tilde{r} = \sqrt{2} N^{1/6} r, \quad (52)$$

as proposed in (49). Furthermore, by matching the two formulas (49) and (51) one finds

$$\nu = -\frac{1}{12}, \quad G(\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} \sqrt{\pi} \text{Ai}(x - \tilde{r}). \quad (53)$$

The next step is then to insert the ansatz (49) into the recursion relation (47) and then perform a large N expansion in the double scaling limit, corresponding to the regime in (48). For this purpose we will also need the large N expansion of the quantities $x_N = \sqrt{2} N^{1/6} (y - \sqrt{2N})$ and $r_N = \sqrt{2} N^{1/6} r$:

$$x_{N+1} = x_N - N^{-1/3} + \mathcal{O}(N^{-1}), \quad r_{N+1} = r_N + \mathcal{O}(N^{-1}). \quad (54)$$

After some straightforward algebra, one finds (remembering that $r = \tilde{r} N^{-1/6}/\sqrt{2}$) the limiting form of the kernel in the double scaling limit, from (37) and (49):

$$K_N(y - r, y - r') \underset{N \rightarrow \infty}{\sim} N^{1/6} \sqrt{2} \frac{G(\tilde{r}, x) \partial_x G(\tilde{r}', x) - \partial_x G(\tilde{r}, x) G(\tilde{r}', x)}{\pi(\tilde{r}' - \tilde{r})}, \quad (55)$$

where $G(\tilde{r}, x)$ is the solution of the Schrödinger equation:

$$-\partial_x^2 G(\tilde{r}, x) + [x + 2q^2(x)] G(\tilde{r}, x) = \tilde{r} G(\tilde{r}, x), \quad (56)$$

with the asymptotic behavior given in Eq. (53) and where $q(x)$ in (56) is the Hastings-McLeod solution of Painlevé II (17).

In view of future purpose, it is also useful to study the asymptotic behavior of $G(\tilde{r}, x)$ when both \tilde{r} and x are large. If one assumes that the large \tilde{r} limit and the large x limit do commute, this asymptotic behavior can be obtained from Eq. (53) where one takes the large \tilde{r} limit. Using the asymptotic behavior of the Airy function for large negative argument, one obtains

$$G(\tilde{r}, x) \sim \frac{1}{\tilde{r}^{1/4}} \sin \left(\frac{2}{3} \tilde{r}^{3/2} - x\sqrt{\tilde{r}} + \frac{\pi}{4} \right), \text{ for } \tilde{r} \gg 1 \text{ \& } x \gg 1, \quad (57)$$

which will be useful in the following to solve the equation for $G(\tilde{r}, x)$ which is the purpose of the next section.

3.2.2 Solution of the Schrödinger equation and the Lax pair of Painlevé XXXIV

We first notice that, if we set $\tilde{r} = 0$ in (56), we see, given the large x behavior (53) that $G(0, x)$ coincides, up to a constant, with $q(x)$ in Eq. (17)

$$G(\tilde{r} = 0, x) = \sqrt{\pi} q(x), \quad (58)$$

which is already an interesting result. The second observation is that the Schrödinger equation (56) has a supersymmetric structure. This property follows from the fact that $q(x)$ is solution of the Painlevé II equation (17), which implies

$$x + 2q^2(x) = \frac{q''(x)}{q(x)} = \frac{d}{dx} \left(\frac{q'(x)}{q(x)} \right) + \left(\frac{q'(x)}{q(x)} \right)^2 = -Q'(x) + Q^2(x), \quad (59)$$

$$\text{with } Q(x) = -\frac{q'(x)}{q(x)}. \quad (60)$$

The next step is to realize that $Q(x) = -q'(x)/q(x)$ can actually be expressed in terms of the Hastings-McLeod solution of the Painlevé II equation with parameter $\alpha = 1/2$, namely (see for instance Ref. [27, 45])

$$Q(x) = 2^{1/3} q_{1/2}(-2^{1/3} x), \quad (61)$$

where $q_{1/2}(s)$ satisfies

$$q_{1/2}''(s) = 2q_{1/2}^3(s) + sq_{1/2}(s) - \frac{1}{2}, \quad (62)$$

with the asymptotic behavior (characterizing the Hastings-McLeod solution in this case):

$$q_{1/2}(s) \underset{s \rightarrow \infty}{\sim} \frac{1}{2s}, \quad (63)$$

while for large negative argument one has

$$q_{1/2}(s) \underset{s \rightarrow -\infty}{\sim} \sqrt{-\frac{s}{2}}. \quad (64)$$

From the definition of $Q(x) = -q'(x)/q(x)$ together with the definition of $q(x)$ (17) it is indeed straightforward to check that $q_{1/2}(s = -2^{1/3}x)$ defined through Eq. (61) satisfies the Painlevé II equation (62) together with the asymptotic behavior in (63). Hence, using Eqs. (59), (61) one can express $G(\tilde{r}, x)$ as

$$G(\tilde{r}, x) = \phi(\tilde{r}, -2^{1/3}x), \text{ where } \partial_s^2 \phi(\tilde{r}, s) - \left[q_{1/2}'(s) + q_{1/2}^2(s) \right] \phi(\tilde{r}, s) = -\frac{\tilde{r}}{2^{2/3}} \phi(\tilde{r}, s), \quad (65)$$

with the appropriate asymptotic behavior of $\phi(\tilde{r}, s)$ deduced from (53):

$$\phi(\tilde{r}, s) \underset{s \rightarrow -\infty}{\sim} \sqrt{\pi} \text{Ai}(-2^{1/3}s - \tilde{r}). \quad (66)$$

The final step is to relate $\phi(\tilde{r}, s)$, which satisfies this supersymmetric Schrödinger equation (65), to a solution of the Lax pair associated to a particular Painlevé XXXIV equation [38]. A similar strategy was used in Ref. [42] to solve a related supersymmetric Schrödinger equation. The relevant Lax pair for our problem is the following one [38, 43]

$$\frac{\partial}{\partial r} \begin{pmatrix} f(r, s) \\ g(r, s) \end{pmatrix} = \mathbf{A} \begin{pmatrix} f(r, s) \\ g(r, s) \end{pmatrix}, \quad \frac{\partial}{\partial s} \begin{pmatrix} f(r, s) \\ g(r, s) \end{pmatrix} = \mathbf{B} \begin{pmatrix} f(r, s) \\ g(r, s) \end{pmatrix}, \quad (67)$$

where \mathbf{A} and \mathbf{B} are 2×2 matrices given by

$$\mathbf{A} = \begin{pmatrix} 2q_{1/2}(s) & 2 + [q_{1/2}^2(s) + q'_{1/2}(s) + \frac{s}{2}] / r \\ -2r - q_{1/2}^2(s) + q'_{1/2}(s) - \frac{s}{2} & -2q_{1/2}(s) \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} q_{1/2}(s) & 1 \\ -r & -q_{1/2}(s) \end{pmatrix}, \quad (68)$$

where the solutions $f(r, s)$ and $g(r, s)$ are characterized by the asymptotic behavior *on the positive real axis* (note in particular that for large negative \tilde{r} , the asymptotic behaviors are different [38], see also below) [38, 43]

$$f(r, s) = r^{-1/4} \sin \left(\frac{4}{3} r^{3/2} + sr^{1/2} + \frac{\pi}{4} \right) + \mathcal{O}(r^{-3/4}) \quad (69a)$$

$$g(r, s) = r^{1/4} \cos \left(\frac{4}{3} r^{3/2} + sr^{1/2} + \frac{\pi}{4} \right) + \mathcal{O}(r^{-1/4}). \quad (69b)$$

The connection between $\phi(\tilde{r}, s)$ and the solution of this Lax system in Eqs. (67, 68) is through the function $f(r, s)$. It is indeed easy to show from Eqs. (67, 68) that $f(r, s)$ satisfies the following Schrödinger equation

$$\partial_s^2 f(r, s) - [q'_{1/2}(s) + q_{1/2}^2(s)] f(r, s) = -r f(r, s), \quad (70)$$

which is exactly similar to the equation satisfied by $\phi(\tilde{r}, s)$ in Eq. (65) with the substitution $r = 2^{-2/3}\tilde{r}$. One can study the large negative s behavior of $f(r, s)$ by plugging the large negative s behavior of $q_{1/2}(s)$ (64) in Eq. (70). This yields:

$$\partial_s^2 f(2^{-2/3}\tilde{r}, s) + \frac{s}{2} f(2^{-2/3}\tilde{r}, s) = -2^{-2/3}\tilde{r} f(2^{-2/3}\tilde{r}, s), \quad \text{when } s \rightarrow -\infty. \quad (71)$$

Hence, one gets that for $s \rightarrow -\infty$,

$$f(2^{-2/3}\tilde{r}, s) \sim a \text{Ai}(-2^{-1/3}s - \tilde{r}) + b \text{Bi}(-2^{-1/3}s - \tilde{r}). \quad (72)$$

Under the (reasonable) assumption that $f(2^{-2/3}\tilde{r}, s)$ remains bounded, one obtains that $b = 0$. On the other hand, the amplitude a can be fixed as we know the large \tilde{r} behavior of $f(2^{-2/3}\tilde{r}, s)$ from Eq. (69). This yields $a = \sqrt{\pi} 2^{1/6}$ and, hence, from Eq. (65) together with (66) one obtains

$$G(\tilde{r}, x) = 2^{-1/6} f(2^{-2/3}\tilde{r}, -2^{1/3}x). \quad (73)$$

We now come back to the kernel $K_N(y - r, y - r')$ in the double scaling limit. From the above relation (73), together with the \mathbf{B} -equation of the Lax pair (67), one can express $\partial_x G(\tilde{r}, x)$ in terms of the functions f and g to obtain (55) in the double scaling limit [we recall that $y = \sqrt{2N} + \frac{x}{\sqrt{2}} N^{-\frac{1}{6}}$ (48)]

$$K_N \left(y - \frac{\tilde{r}}{\sqrt{2}} N^{-1/6}, y - \frac{\tilde{r}'}{\sqrt{2}} N^{-1/6} \right) \underset{N \rightarrow \infty}{\sim} \frac{N^{1/6} \sqrt{2} f(2^{-2/3}\tilde{r}, -2^{1/3}x) g(2^{-2/3}\tilde{r}', -2^{1/3}x) - g(2^{-2/3}\tilde{r}, -2^{1/3}x) f(2^{-2/3}\tilde{r}', -2^{1/3}x)}{\pi(\tilde{r} - \tilde{r}')} . \quad (74)$$

In view of this expression (74), it is natural to introduce the following functions $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ such that

$$f(2^{-2/3}\tilde{r}, -2^{1/3}x) = 2^{1/3} \tilde{f}(\tilde{r}, x), \quad g(2^{-2/3}\tilde{r}, -2^{1/3}x) = \tilde{g}(\tilde{r}, x), \quad (75)$$

from which it follows, using Eq. (73), that

$$G(\tilde{r}, x) = 2^{1/6} \tilde{f}(\tilde{r}, x). \quad (76)$$

Finally, one has

$$K_N \left(y - \frac{\tilde{r}}{\sqrt{2}} N^{-1/6}, y - \frac{\tilde{r}'}{\sqrt{2}} N^{-1/6} \right) \underset{N \rightarrow \infty}{\sim} N^{1/6} 2^{5/6} \frac{\tilde{f}(\tilde{r}, x) \tilde{g}(\tilde{r}', x) - \tilde{f}(\tilde{r}', x) \tilde{g}(\tilde{r}, x)}{\pi(\tilde{r} - \tilde{r}')} . \quad (77)$$

From the Lax pair (67), (68), together with some identities satisfied by $q_{1/2}(x)$, we show in Appendix A that $\tilde{f}(r, s)$ and $\tilde{g}(r, s)$ are solutions of a system of differential equations which can be expressed in terms of $q(x)$, the Hastings-McLeod solution of the Painlevé II equation with $\alpha = 0$ (17):

$$\frac{\partial}{\partial \tilde{r}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{A}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad \frac{\partial}{\partial x} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix} = \tilde{\mathbf{B}} \begin{pmatrix} \tilde{f}(\tilde{r}, x) \\ \tilde{g}(\tilde{r}, x) \end{pmatrix}, \quad (78)$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are 2×2 matrices given by

$$\tilde{\mathbf{A}} = \begin{pmatrix} -\frac{q'(x)}{q(x)} & 1 + q^2(x)/\tilde{r} \\ -\tilde{r} - \frac{\int_x^\infty q^2(u) du}{q^2(x)} & \frac{q'(x)}{q(x)} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \frac{q'(x)}{q(x)} & -1 \\ \tilde{r} & -\frac{q'(x)}{q(x)} \end{pmatrix}, \quad (79)$$

where the solutions $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ are characterized by the asymptotic behavior inherited from (69)

$$\tilde{f}(\tilde{r}, x) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-1/6} \tilde{r}^{-1/4} \sin \left(\frac{2}{3} \tilde{r}^{3/2} - x\sqrt{\tilde{r}} + \frac{\pi}{4} \right) + \mathcal{O}(\tilde{r}^{-3/4}), \quad (80a)$$

$$\tilde{g}(\tilde{r}, x) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-1/6} \tilde{r}^{1/4} \cos \left(\frac{2}{3} \tilde{r}^{3/2} - x\sqrt{\tilde{r}} + \frac{\pi}{4} \right) + \mathcal{O}(\tilde{r}^{-1/4}). \quad (80b)$$

From this expression (77) one can compute the DOS $\rho_{\text{DOS}}(r, N)$ for $r \sim \mathcal{O}(N^{-1/6})$. The above asymptotics (80) will be useful to study the asymptotic behavior of $\rho_{\text{DOS}}(r, N)$ for large r .

3.2.3 Limiting expression of the DOS in the double scaling limit

For the purpose of the computation of $\rho_{\text{DOS}}(r, N)$ given in (34), we need to compute $K_N(y - \tilde{r}N^{-1/6}/\sqrt{2}, y)$ which is given, from Eq. (77), by

$$K_N(y - r, y) \underset{N \rightarrow \infty}{\sim} N^{1/6} 2^{5/6} \frac{\tilde{f}(\tilde{r}, x) \tilde{g}(0, x) - \tilde{g}(\tilde{r}, x) \tilde{f}(0, x)}{\pi \tilde{r}}. \quad (81)$$

Using that $\tilde{g}(0, x) = 0$ (see Appendix B), together with $\tilde{f}(0, x) = 2^{-1/6} \sqrt{\pi} q(x)$ [see Eqs. (58), (76)], one obtains:

$$K_N \left(y - \frac{\tilde{r}}{\sqrt{2}} N^{-1/6}, y \right) \underset{N \rightarrow \infty}{\sim} -N^{1/6} 2^{2/3} \frac{\tilde{g}(\tilde{r}, x) q(x)}{\sqrt{\pi} \tilde{r}}. \quad (82)$$

On the other hand, we also need the kernel at coinciding point (34). It can easily be obtained from Eq. (77) using l'Hospital's rule

$$K_N(y - r, y - r) \underset{N \rightarrow \infty}{\sim} \frac{2^{5/6} N^{1/6}}{\pi} \left[\partial_{\tilde{r}} \tilde{f}(\tilde{r}, x) \tilde{g}(\tilde{r}, x) - \tilde{f}(\tilde{r}, x) \partial_{\tilde{r}} \tilde{g}(\tilde{r}, x) \right] \quad (83)$$

$$= \frac{2^{5/6} N^{1/6}}{\pi} \left[\left(\tilde{r} + \frac{\int_x^\infty q^2(u) du}{q^2(x)} \right) \tilde{f}^2 - 2 \frac{q'(x)}{q(x)} \tilde{f} \tilde{g} + \left(1 + \frac{q^2(x)}{\tilde{r}} \right) \tilde{g}^2 \right]. \quad (84)$$

In particular, using that $\tilde{g}(\tilde{r}, x) = \mathcal{O}(r)$, when $r \rightarrow 0$ (see Appendix B) and $\tilde{f}(0, x) = 2^{-1/6} \sqrt{\pi} q(x)$ one finds from (83) that

$$K_N(y, y) \underset{N \rightarrow \infty}{\sim} \sqrt{2} N^{1/6} \int_x^\infty q^2(u) du . \quad (85)$$

One can check that this relation (85) is consistent with the identity shown above (42). Indeed, one has

$$\log F_N(y) \underset{N \rightarrow \infty}{\rightarrow} \log \mathcal{F}_2(\sqrt{2} N^{1/6} (y - \sqrt{2N})) , \quad (86)$$

where $\mathcal{F}_2(x)$ is the Tracy-Widom distribution for $\beta = 2$, given by [20]

$$\log \mathcal{F}_2(\sqrt{2} N^{1/6} (y - \sqrt{2N})) = - \int_x^\infty (u - x) q^2(u) du , \quad (87)$$

hence

$$\frac{\partial}{\partial y} \log F_N(y) = \sqrt{2} N^{1/6} \int_x^\infty q^2(u) du . \quad (88)$$

One finds finally, that

$$\rho_{\text{DOS}}(r, N) = N^{-5/6} \sqrt{2} \tilde{\rho}_{\text{edge}}(\sqrt{2} N^{1/6} r) , \quad (89a)$$

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^\infty \left[R \left(\left(\tilde{r} + \frac{R}{q^2} \right) \tilde{f}^2 - 2 \frac{q'}{q} \tilde{f} \tilde{g} + \left(1 + \frac{q^2}{\tilde{r}} \right) \tilde{g}^2 \right) - \frac{1}{\tilde{r}^2} q^2 \tilde{g}^2 \right] \mathcal{F}_2 dx , \quad (89b)$$

where we have used the shorthand notations $q = q(x)$, $R = R(x)$, $\mathcal{F}_2 = \mathcal{F}_2(x)$, $\tilde{f} = \tilde{f}(\tilde{r}, x)$ and $\tilde{g} = \tilde{g}(\tilde{r}, x)$ as well as

$$R(x) = \int_x^\infty q^2(u) du . \quad (90)$$

This expression (89) is still a bit cumbersome but we can further simplify it, by using the remarkable identity

$$\partial_x \left(\left(\tilde{r} + \frac{R}{q^2} \right) \tilde{f}^2 - 2 \frac{q'}{q} \tilde{f} \tilde{g} + \left(1 + \frac{q^2}{\tilde{r}} \right) \tilde{g}^2 \right) = -\tilde{f}^2 , \quad (91)$$

which can be checked using the fact that \tilde{f} and \tilde{g} are solutions of the Lax system (78). Hence using this identity in Eq. (91) together with (90), one can write $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in Eq. (89) as

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^\infty \left[R \int_x^\infty \tilde{f}^2(\tilde{r}, u) du - q^2 \frac{\tilde{g}}{\tilde{r}^2} \right] \mathcal{F}_2(x) dx . \quad (92)$$

Besides, from the equation satisfied by $\partial_x \tilde{g}(\tilde{r}, x)$ (78, 79), one has

$$\partial_x (q(x) \tilde{g}(\tilde{r}, x)) = \tilde{r} q(x) \tilde{f}(\tilde{r}, x) \implies q(x) \tilde{g}(\tilde{r}, x) = -\tilde{r} \int_x^\infty q(u) \tilde{f}(\tilde{r}, u) du , \quad (93)$$

where we have used that $g(\tilde{r}, x) \rightarrow 0$ when $x \rightarrow \infty$. Performing an integration by part in (92), and using that $R = \mathcal{F}_2' / \mathcal{F}_2$ together with (93), we obtain

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^\infty \left[\tilde{f}^2 - \frac{q^2}{\tilde{r}^2} \tilde{g}^2 \right] \mathcal{F}_2 dx = \frac{2^{1/3}}{\pi} \int_{-\infty}^\infty \left[\tilde{f}^2 - \left(\int_x^\infty q \tilde{f} du \right)^2 \right] \mathcal{F}_2 dx , \quad (94)$$

where we have used the notation $q \tilde{f} = q(u) \tilde{f}(\tilde{r}, u)$. We recall that $\mathcal{F}_2(x)$ is the Tracy-Widom distribution associated to GUE, which is given by

$$\mathcal{F}_2(x) = \exp \left[- \int_x^\infty (u - x) q^2(u) du \right] , \quad (95)$$

while $\tilde{f}(\tilde{r}, x)$ satisfies

$$\partial_x^2 \tilde{f}(\tilde{r}, x) - [x + 2q^2(x)] \tilde{f}(\tilde{r}, x) = -\tilde{r} \tilde{f}(\tilde{r}, x), \quad \tilde{f}(\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} 2^{-1/6} \sqrt{\pi} \text{Ai}(x - \tilde{r}), \quad (96)$$

and $q(x)$ is the Hastings-McLeod solution of the Painlevé II equation (17). The expression given in Eq. (94) is one of the main results of the present paper. We now analyze the asymptotic behaviors of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ for small and large argument.

3.2.4 Asymptotic behavior for large \tilde{r}

To analyze the large \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$, it turns out that it is much easier to use the expression in Eq. (89). To this purpose we use the large \tilde{r} behavior of \tilde{f} and \tilde{g} in Eq. (80). One finds that the leading terms in the integrand in (89) are

$$\left[R \left(\left(\tilde{r} + \frac{R}{q^2} \right) \tilde{f}^2 - 2 \frac{q'}{q} \tilde{f} \tilde{g} + \left(1 + \frac{q^2}{\tilde{r}} \right) \tilde{g}^2 \right) - \frac{1}{\tilde{r}^2} q^2 \tilde{g}^2 \right] \sim R(\tilde{r} \tilde{f}^2 + \tilde{g}^2) \quad (97a)$$

$$\sim 2^{-1/3} \sqrt{\tilde{r}} R. \quad (97b)$$

Plugging this expansion (97) into the expression for $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in Eq. (89) and using $R(x) \mathcal{F}_2(x) = d\mathcal{F}_2(x)/dx$ one obtains straightforwardly from (89) the leading asymptotic behavior

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \frac{\sqrt{\tilde{r}}}{\pi}, \quad (98)$$

which yields the second line of Eq. (22). As explained above, below Eq. (22), this asymptotic behavior (98) ensures a perfect matching between the edge regime of $\rho_{\text{DOS}}(r, N)$ when $r = \mathcal{O}(N^{-1/6})$, described by $\tilde{\rho}_{\text{edge}}$ (89, 94), and the bulk regime when $r = \mathcal{O}(\sqrt{N})$, described by a shifted Wigner semi-circle law in Eq. (14). Note that one can show, and this is confirmed by a numerical evaluation of our formula for $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in Eq. (94) (see section 5), that the first corrections to the leading behavior in Eq. (98) are of order $\mathcal{O}(\tilde{r}^{-1/2})$.

3.2.5 Asymptotic behavior for small \tilde{r}

The study of the small \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$, which is more conveniently performed on Eq. (94), requires the analysis of the Lax system (67) when $\tilde{r} \rightarrow 0$. Here we present the analysis of the small \tilde{r} expansion at lowest order, which requires the expansion of $\tilde{f}(\tilde{r}, x)$ up to order $\mathcal{O}(\tilde{r}^3)$ (see Appendix B):

$$\tilde{f}(\tilde{r}, x) = \tilde{f}(0, x) + \tilde{r} \tilde{f}_1(x) + \tilde{r}^2 \tilde{f}_2(x) + \mathcal{O}(\tilde{r}^3), \quad (99)$$

where the functions $\tilde{f}_k(x)$ can be computed explicitly along the lines explained in Appendix B yielding:

$$\tilde{f}(0, x) = 2^{-1/6} \sqrt{\pi} q, \quad (100a)$$

$$\tilde{f}_1(x) = -2^{-1/6} \sqrt{\pi} [q' + qR], \quad (100b)$$

$$\tilde{f}_2(x) = 2^{-7/6} \sqrt{\pi} \left[\frac{q'^2}{q} + q'R - \frac{R}{q} - \frac{1}{2} q^3 + \frac{q}{2} R^2 \right]. \quad (100c)$$

One can then insert this expansion (99, 100) into Eq. (94) to obtain

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \int_{-\infty}^{\infty} (q^2(x) - R^2(x)) \mathcal{F}_2(x) dx \quad (101a)$$

$$+ \tilde{r} \frac{2^{1/6}}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(q(x) \tilde{f}_1(x) - R(x) \int_x^{\infty} q(u) \tilde{f}_1(u) du \right) \mathcal{F}_2(x) dx \quad (101b)$$

$$+ \tilde{r}^2 \left[\frac{2^{1/6}}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(q(x) \tilde{f}_2(x) - R(x) \int_x^{\infty} q(u) \tilde{f}_2(u) du \right) \mathcal{F}_2(x) dx \right. \quad (101c)$$

$$\left. + \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left(\tilde{f}_1^2(x) - \left(\int_x^{\infty} q(u) \tilde{f}_1(u) du \right)^2 \right) \mathcal{F}_2(x) dx + \mathcal{O}(\tilde{r}^3) \right]. \quad (101d)$$

By performing integration by parts, using $R(x) = \mathcal{F}'_2(x)/\mathcal{F}_2(x)$, one finds that the three first terms (101a)-(101c) in this expansion all vanish. One finds finally

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \tilde{r}^2 \int_{-\infty}^{\infty} \left[(q' + qR)^2 - \frac{1}{4}(q^2 - R^2)^2 \right] \mathcal{F}_2 dx + \mathcal{O}(\tilde{r}^3). \quad (102)$$

After multiple integration by parts one can compute exactly this integral over x (see Appendix B). This yields the very simple result

$$a_2 = \int_{-\infty}^{\infty} \left[(q' + qR)^2 - \frac{1}{4}(q^2 - R^2)^2 \right] \mathcal{F}_2 dx = \frac{1}{2}, \quad (103)$$

so that, to leading order, $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \tilde{r}^2/2$. From the expression for $\tilde{\rho}_{\text{edge}}(\tilde{r})$ obtained in (94) it is possible to go beyond the leading order, as explained in Appendix B. In fact, to compute $\tilde{\rho}_{\text{edge}}(\tilde{r})$ up to order $\mathcal{O}(\tilde{r}^4)$ one can show that this is sufficient to know $\tilde{f}(\tilde{r}, x)$ only up to order $\mathcal{O}(\tilde{r}^2)$, as given in Eq. (99). We then obtain explicitly that the term of order $\mathcal{O}(\tilde{r}^3)$ vanishes, yielding $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \tilde{r}^2/2 + a_4\tilde{r}^4$ where a_4 is given below (105).

Finally, the asymptotic behaviors of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ can be summarized, as announced in Eq. (22), as

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \begin{cases} \frac{1}{2} \tilde{r}^2 + a_4 \tilde{r}^4, & \tilde{r} \rightarrow 0 \\ \frac{\sqrt{\tilde{r}}}{\pi}, & \tilde{r} \rightarrow \infty \end{cases}, \quad (104)$$

where the amplitude a_4 is given by the integral

$$a_4 = \frac{1}{2} \int_{-\infty}^{\infty} \left[H(x) + \frac{1}{2}(T^2(x) - H^2(x)) \right] \mathcal{F}_2(x) dx, \quad (105)$$

$$H(x) = -\frac{1}{2}q^2R + \frac{1}{6}R^3 + \int_x^{\infty} (q^4(u) + uq^2(u))du, \quad T(x) = \frac{H'(x)}{q(x)}.$$

This integral (105) can be evaluated numerically [44] to yield $a_4 = -0.393575\dots$. The quadratic behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ for small \tilde{r} can be qualitatively understood as it is related to the probability that two eigenvalues become extremely close to each other – namely the first one and the second one [see Eq. (23)]. From the joint PDF of the eigenvalues (2), this probability vanishes quadratically for GUE. More generally, for the Gaussian β -ensemble, one thus expects that $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta$, while there is no obvious simple argument to compute the constant a_β , even for $\beta = 2$ where $a_2 = 1/2$, which arises in our calculation in a non trivial way. On the other hand, if we assume a smooth matching between the edge region and the bulk region described by the shifted Wigner semi-circle law (46), then the large \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$, to leading order, immediately follows: $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \sqrt{\tilde{r}}/\pi$ to match with the small argument of Eq. (46). It is natural to assume that the matching holds for all values of β . Hence one expects the following asymptotic behaviors

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \begin{cases} a_\beta \tilde{r}^\beta + o(\tilde{r}^\beta), & \tilde{r} \rightarrow 0 \\ \frac{\sqrt{\tilde{r}}}{\pi} + o(\tilde{r}^{1/2}), & \tilde{r} \rightarrow \infty \end{cases}, \quad (106)$$

for the Gaussian β -ensemble.

We now show that the double scaling analysis carried out to compute $\tilde{\rho}_{\text{edge}}(\tilde{r})$ can be straightforwardly adapted to obtain a relatively simpler formula (compared to the formula obtained by WBF in Ref. [27] and given in Appendix D) for the distribution of the typical fluctuations of the first (scaled) gap in the large N limit.

4 Application to the distribution of the gap between the two largest eigenvalues

We now focus on the first gap between the two largest eigenvalues. If one denotes by $\lambda_{\max} = \Lambda_{1,N} \geq \Lambda_{2,N} \geq \dots \geq \Lambda_{N,N}$ we compute the PDF $p_{\text{GAP}}(r, N)$ of the random variable $d_{1,N} = \Lambda_{1,N} - \Lambda_{2,N}$ (see Fig. 1).

4.1 Exact expression for any finite N and large N expansion

An exact expression for $p_{\text{GAP}}(r, N)$ can be easily written from the joint PDF of the N eigenvalues in Eq. (2). For this purpose, we notice that the probability $p_{\text{GAP}}(r, N)dr = \text{Pr.}(\Lambda_{1,N} - \Lambda_{2,N} \in [r, r+dr])$ is obtained by integrating over the value $y \in (-\infty, +\infty)$ of the second eigenvalue $\Lambda_{2,N}$, while $\Lambda_{1,N} = \lambda_{\max}$ is fixed to be $y + r$. This means in particular that there are $(N - 2)$ eigenvalues whose value is less than y . Hence:

$$p_{\text{GAP}}(r, N) = N(N-1) \int_{-\infty}^{+\infty} dy \int_{-\infty}^y d\lambda_1 \int_{-\infty}^y d\lambda_2 \dots \int_{-\infty}^y d\lambda_{N-2} P_{\text{joint}}(\lambda_1, \dots, \lambda_{N-2}, y, y+r), \quad (107)$$

where the factor $N(N - 1)$ in Eq. (107) comes from the $N(N - 1)$ ways of choosing the pair of the two largest eigenvalues $(\Lambda_{1,N}, \Lambda_{2,N})$ among N . By comparing with the exact expression in (27), using the invariance of the joint PDF (2) under permutation of two eigenvalues, one obtains that $p_{\text{GAP}}(r, N) = (N - 1)\rho_{\text{DOS}}(-r, N)$, as announced in Eq. (23) ⁴.

From this identity, we can now use the previous analysis to study the typical fluctuations of the first gap $d_{1,N}$, i.e. for $d_{1,N} = \mathcal{O}(N^{-1/6})$. This yields immediately the expression announced in Eq. (15):

$$p_{\text{GAP}}(r, N) = \sqrt{2}N^{1/6}\tilde{p}_{\text{typ}}\left(r\sqrt{2}N^{1/6}\right), \quad (108a)$$

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \left(\int_x^{\infty} q(u)\tilde{f}(-\tilde{r}, u)du \right)^2 \right] \mathcal{F}_2(x) dx, \quad (108b)$$

where $\tilde{f}(-\tilde{r}, x)$ is solution of the Schrödinger equation:

$$\partial_x^2 \tilde{f}(-\tilde{r}, x) - [x + 2q^2(x)]\tilde{f}(-\tilde{r}, x) = \tilde{r}\tilde{f}(\tilde{r}, x), \quad \tilde{f}(-\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} 2^{-1/6}\sqrt{\pi}\text{Ai}(x + \tilde{r}), \quad \tilde{r} > 0. \quad (109)$$

One can again show that $\tilde{f}(-\tilde{r}, x)$ is the first component of the $2d$ -vector which is solution of the Lax system in Eq. (78) where the matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are given in Eq. (79) with the substitution $\tilde{r} \rightarrow -\tilde{r}$. This change of sign has drastic consequences on the asymptotic behaviors of $\tilde{f}(-\tilde{r}, x)$ and $\tilde{g}(-\tilde{r}, x)$. Instead of an oscillating behavior as in Eq. (80) they are characterized in this case by an exponential decay. This is also clear on the above Schrödinger equation (109). One has indeed [38]

$$\tilde{f}(-\tilde{r}, s) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-7/6}\tilde{r}^{-1/4} \exp\left(-\frac{2}{3}\tilde{r}^{3/2} - s\sqrt{\tilde{r}}\right) \left(1 + \mathcal{O}(\tilde{r}^{-1/2})\right), \quad (110a)$$

$$\tilde{g}(-\tilde{r}, s) \underset{\tilde{r} \rightarrow \infty}{\sim} 2^{-7/6}\tilde{r}^{1/4} \exp\left(-\frac{2}{3}\tilde{r}^{3/2} - s\sqrt{\tilde{r}}\right) \left(1 + \mathcal{O}(\tilde{r}^{-1/2})\right), \quad (110b)$$

which are essential to study the large \tilde{r} asymptotics of $\tilde{p}_{\text{typ}}(\tilde{r})$.

⁴ See the remark below Eq. (37) concerning the definition of $\rho_{\text{DOS}}(r, N)$ for negative r .

4.2 Asymptotic behavior of $\tilde{p}_{\text{typ}}(\tilde{r})$ for small argument

The small \tilde{r} asymptotic of $\tilde{p}_{\text{typ}}(\tilde{r})$ is exactly similar to the one performed before for $\tilde{\rho}_{\text{edge}}(\tilde{r})$ as one can argue that it involves only even powers of \tilde{r} . Indeed, for small \tilde{r} , the contribution to $\tilde{\rho}_{\text{edge}}(\tilde{r})$ come only from the gap between the two largest eigenvalues because the others (the third, the fourth etc. eigenvalues) are too far. This implies that $\tilde{\rho}_{\text{edge}}(\tilde{r}) \simeq \tilde{p}_{\text{typ}}(\tilde{r})$. But on the other hand the identity in (23) implies $\tilde{p}_{\text{typ}}(\tilde{r}) \simeq \tilde{\rho}_{\text{edge}}(-\tilde{r})$. Hence, if one assumes that $\tilde{\rho}_{\text{edge}}(\tilde{r})$ is analytic in $\tilde{r} = 0$, one expects that the small \tilde{r} expansion of $\tilde{p}_{\text{typ}}(\tilde{r})$ [or equivalently the one of $\tilde{\rho}_{\text{edge}}(\tilde{r})$] only involves even powers of \tilde{r} . In particular, the two first terms are [see Eq. (104)]

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{1}{2}\tilde{r}^2 + a_4\tilde{r}^4 + \mathcal{O}(\tilde{r}^6), \quad (111)$$

as announced in the introduction in Eq. (26). The quadratic behavior in (111) can be understood physically as coming from the short distance repulsion between eigenvalues, which comes from the Vandermonde term $\prod_{i,j}(\lambda_i - \lambda_j)^2$ in the joint PDF for GUE (2). Similarly to the DOS (106), one thus expects that, for the Gaussian β -ensemble, one has $\tilde{p}_{\text{typ}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta$ (see also [8] in the case of Gaussian Orthogonal Ensemble, corresponding to $\beta = 1$).

4.3 Asymptotic behavior of $\tilde{p}_{\text{typ}}(\tilde{r})$ for large argument

The asymptotic analysis of $\tilde{p}_{\text{typ}}(\tilde{r})$ for large \tilde{r} is quite different from the one performed above for $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in section 3.2.5. This difference, as mentioned above, is due to the qualitatively different behaviors of $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ in Eq. (80) and $\tilde{f}(-\tilde{r}, x)$ and $\tilde{g}(-\tilde{r}, x)$ in Eq. (110) when \tilde{r} is large. To perform this large \tilde{r} analysis, it is more convenient to use the following expression in terms of $\tilde{f}(-\tilde{r}, x)$ and $\tilde{g}(-\tilde{r}, x)$ [as in the first expression given in Eq. (94) with the substitution $\tilde{r} \rightarrow -\tilde{r}$]:

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \frac{1}{r^2} q^2(x) \tilde{g}^2(-\tilde{r}, x) \right] \mathcal{F}_2(x) dx. \quad (112)$$

By plugging the large \tilde{r} behaviors of $\tilde{f}(-\tilde{r}, x)$ and $\tilde{g}(-\tilde{r}, x)$ given in Eq. (110) in the above expression (112), one obtains

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{1}{4\pi} \frac{1}{\tilde{r}^{1/2}} e^{-\frac{4}{3}\tilde{r}^{3/2}} I(\tilde{r}) \left(1 + \mathcal{O}(\tilde{r}^{-1/2}) \right), \quad I(\tilde{r}) = \int_{-\infty}^{\infty} e^{-2x\sqrt{\tilde{r}}} \mathcal{F}_2(x) dx. \quad (113)$$

The leading corrections, of order $\mathcal{O}(\tilde{r}^{-1/2})$, in Eq. (113) come from the leading corrections to the large \tilde{r} behavior of $\tilde{f}(-\tilde{r}, x)$ in Eq. (110) (these terms are analyzed in detail in Appendix C). We thus focus now on the large \tilde{r} expansion of $I(\tilde{r})$ in (113). To analyze it, we first split the integral over x in (113) into two parts: one on \mathbb{R}^- and another one on \mathbb{R}^+ :

$$I(\tilde{r}) = I_+(\tilde{r}) + I_-(\tilde{r}), \quad (114a)$$

$$I_+(\tilde{r}) = \int_0^{\infty} e^{-2x\sqrt{\tilde{r}}} \mathcal{F}_2(x) dx, \quad I_-(\tilde{r}) = \int_{-\infty}^0 e^{-2x\sqrt{\tilde{r}}} \mathcal{F}_2(x) dx. \quad (114b)$$

The analysis of $I_+(\tilde{r})$ is easily done, using that $\mathcal{F}_2(x) \leq 1$, for all x , which implies

$$0 \leq I_+(\tilde{r}) \leq \int_0^{\infty} e^{-2x\sqrt{\tilde{r}}} dx = \frac{1}{2\sqrt{\tilde{r}}}, \quad (115)$$

such that $I_+(\tilde{r}) \rightarrow 0$ when $\tilde{r} \rightarrow \infty$.

The analysis of $I_-(\tilde{r})$ is quite different. Indeed, because of the exponential term $e^{-2x\sqrt{\tilde{r}}}$, with $x < 0$, the integral over x in $I_-(\tilde{r})$ is dominated, when $\tilde{r} \gg 1$, by the region of large negative x . Hence for large \tilde{r} , one can replace, in $I_-(\tilde{r})$, $\mathcal{F}_2(x)$ by its asymptotic behavior for large negative argument. It reads [46]:

$$\mathcal{F}_2(x) = \tau_2 |x|^{-1/8} e^{-\frac{1}{12}|x|^3} (1 + \mathcal{O}(|x|^{-3})) , \quad \tau_2 = 2^{1/24} e^{\zeta'(-1)}. \quad (116)$$

By plugging the asymptotic behavior of $\mathcal{F}_2(x)$ (116) into the integral defining $I_-(\tilde{r})$ in Eq. (114) we see that the integrand is of the form $\exp(2|x|\sqrt{\tilde{r}} - \frac{1}{12}|x|^3)$: hence for large \tilde{r} , this integral can be evaluated by the saddle point method. Besides, by balancing the two terms in the argument of the exponential $|x|\sqrt{\tilde{r}}$ and $|x|^3$, we obtain that the saddle point x^* is of order $\mathcal{O}(\tilde{r}^{1/4})$ – and negative. This suggests to perform the change of variable $x = -u\tilde{r}^{1/4}$ in Eq. (114):

$$I_-(\tilde{r}) = \tilde{r}^{1/4} \int_0^\infty e^{2u\tilde{r}^{3/4}} \mathcal{F}_2(-u\tilde{r}^{1/4}) du \quad (117a)$$

$$= \tau_2 \tilde{r}^{7/32} \int_0^\infty \frac{du}{u^{1/8}} e^{\tilde{r}^{3/4} \phi(u)} \left(1 + \mathcal{O}(\tilde{r}^{-3/4})\right), \quad \phi(u) = 2u - \frac{u^3}{12}, \quad (117b)$$

where the terms of order $\mathcal{O}(\tilde{r}^{-3/4})$ coming from the subleading corrections to $\mathcal{F}_2(x)$ in Eq. (116) are studied in detail in Appendix C. It is easy to see that the function $\phi(u)$ in Eq. (117) admits a maximum in $u^* = 2\sqrt{2}$, where $\phi(u^*) = 8\sqrt{2}/3$ so that the integral over u in (117) can be evaluated using the saddle point method to yield

$$I_-(\tilde{r}) = 2^{5/48} \sqrt{\pi} e^{\zeta'(-1)} \tilde{r}^{-\frac{5}{32}} \exp\left(\frac{8}{3} \sqrt{2} \tilde{r}^{3/4}\right) \left(1 + \mathcal{O}(\tilde{r}^{-3/4})\right). \quad (118)$$

Using that $I(\tilde{r}) \sim I_-(\tilde{r})$ [see Eqs. (114, 115)] together with the analysis of the first corrections to this leading behavior (118) which is performed in Appendix C, we obtain finally the result :

$$\tilde{p}_{\text{typ}}(\tilde{r}) = A \exp\left(-\frac{4}{3} \tilde{r}^{3/2} + \frac{8}{3} \sqrt{2} \tilde{r}^{3/4}\right) \tilde{r}^{-21/32} \left(1 - \frac{1405\sqrt{2}}{1536} \tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2})\right), \quad (119)$$

with $A = 2^{-91/48} e^{\zeta'(-1)} / \sqrt{\pi}$, as announced in the second line of Eq. (26).

The asymptotic behavior of $\tilde{p}_{\text{typ}}(\tilde{r})$ to leading order, $\tilde{p}_{\text{typ}}(\tilde{r}) \sim \exp(-\frac{4}{3} \tilde{r}^{3/2})$ turns out to be the same as the right tail of the Tracy-Widom distribution, \mathcal{F}_2 , in Eq. (16). This can be understood via the following heuristic argument (see also [27]): for large separation \tilde{r} the two first eigenvalues $\Lambda_{2,N}$ and $\Lambda_{1,N} = \lambda_{\text{max}}$ become statistically independent:

$$\begin{aligned} \tilde{p}_{\text{gap}}(\tilde{r}) &= \int_{-\infty}^\infty \text{Pr}[\Lambda_{2,N} = \lambda, \lambda_{\text{max}} = \lambda + \tilde{r}] d\lambda \underset{r \rightarrow \infty}{\sim} \int_{-\infty}^\infty \text{Pr}[\Lambda_{2,N} = \lambda] \text{Pr}[\lambda_{\text{max}} = \lambda + \tilde{r}] d\lambda \\ &\underset{r \rightarrow \infty}{\sim} \text{Pr}[\lambda_{\text{max}} = \tilde{r}], \end{aligned} \quad (120)$$

which thus naturally yields the right tail of the TW distribution. Note this heuristic argument (120) can be adapted to the Gaussian β -ensemble, for which the TW- β distribution behaves for large argument like $\mathcal{F}'_\beta(x) \sim \exp(-\frac{2\beta}{3} x^{3/2})$. Hence, reminding the small \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ studied in section 4.2, one expects, for the Gaussian β -ensemble:

$$\tilde{p}_{\text{typ}}(\tilde{r}) \sim \begin{cases} a_\beta \tilde{r}^\beta + o(\tilde{r}^\beta) & \tilde{r} \rightarrow 0 \\ \exp\left(-\frac{2\beta}{3} \tilde{r}^{3/2} + o(\tilde{r}^{3/2})\right) & \tilde{r} \rightarrow +\infty, \end{cases} \quad (121)$$

while the computation of a_β as well as the subleading corrections, both for small and large arguments, for any β remains challenging.

5 Numerical evaluation of the limiting scaling functions $\tilde{\rho}_{\text{edge}}(\tilde{r})$ and $\tilde{p}_{\text{typ}}(\tilde{r})$

In this section, we provide a numerical evaluation of the limiting functions $\tilde{\rho}_{\text{edge}}(\tilde{r})$ and $\tilde{p}_{\text{typ}}(\tilde{r})$ given respectively in Eq. (94) and Eq. (108). To evaluate numerically these formula, we need to compute numerically the solution of the Lax pair, (\tilde{f}, \tilde{g}) . We found that the easiest way to evaluate them is to

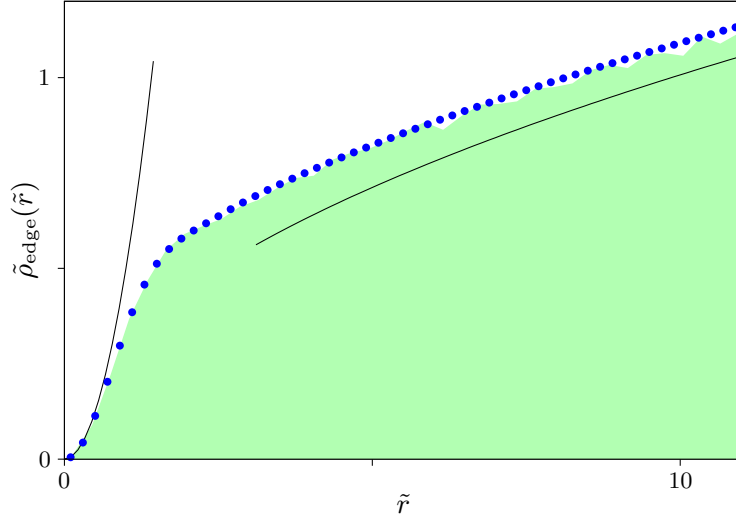


Fig. 4 Blue dots: numerical evaluation of our formula (94) for $\tilde{\rho}_{\text{edge}}(\tilde{r})$, green filling: numerical result with 2×10^5 realizations for a 1000×1000 GUE matrix, black dashed curve: asymptotic behaviors given in Eq. (22).

solve numerically (with Mathematica) the following system of coupled differential equations satisfied by $\tilde{f}(\tilde{r}, x)$ (96) and $\tilde{g}(\tilde{r}, x)$ (93), \tilde{r} being a parameter:

$$\begin{cases} \partial_x^2 \tilde{f}(\tilde{r}, x) = (x + 2q^2 - \tilde{r})\tilde{f}(\tilde{r}, x) & , \quad \tilde{f}(\tilde{r}, x) \underset{x \gg 1}{\approx} 2^{-1/6} \sqrt{\pi} \text{Ai}(x - \tilde{r}) , \\ \partial_x [q(x)\tilde{g}(\tilde{r}, x)] = \tilde{r}q(x)\tilde{f}(\tilde{r}, x) & , \quad \tilde{g}(\tilde{r}, x) \underset{x \gg 1}{\rightarrow} 0 , \end{cases} \quad (122)$$

where q is the Hasting McLeod solution of Painlevé II (17) whose numerical estimation is based on M. Prähofer's tables [44]. Once $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ are evaluated numerically, we can then compute the distributions $\tilde{\rho}_{\text{edge}}(\tilde{r})$ and $\tilde{p}_{\text{typ}}(\tilde{r})$ by evaluating numerically the integrals in Eq. (94) and Eq. (108).

In Fig. 4 we show a plot of the scaling function $\tilde{\rho}_{\text{edge}}(\tilde{r})$, evaluated numerically (the blue dots), as a function of \tilde{r} . The black dashed curves indicate the leading behavior for small \tilde{r} , $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \tilde{r}^2/2$ [see Eq. (104)] and the leading large \tilde{r} behavior, $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \sqrt{\tilde{r}}/\pi$. A numerical estimation of $\tilde{\rho}_{\text{edge}}(\tilde{r}) - \sqrt{\tilde{r}}/\pi$ indicates that this quantity vanishes like $\tilde{r}^{-1/2}$ for large \tilde{r} . Finally, the green filling in Fig. 4 shows a numerical estimation of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ obtained by sampling 2×10^5 independent random Hermitian matrices of size 1000×1000 . The agreement with the numerical evaluation of our formula (94) is quite good, which provides a good check of this formula (94). One should however notice that we have not attempted to evaluate the precision of our numerical procedure to evaluate numerically this formula (94).

We have also evaluated numerically our formula for $\tilde{p}_{\text{typ}}(\tilde{r})$, given in Eq. (108). These numerical values correspond to the red dots shown in Fig. 5. We have compared these values with the numerical evaluation of $\tilde{p}_{\text{typ}}(\tilde{r})$ provided by WBF [27] (the black curve in Fig. 5) – this PDF is actually tabulated in Table 2 of Ref. [27]. Note that these numerical estimates have been obtained by evaluating numerically the expression of $\tilde{p}_{\text{typ}}(\tilde{r})$ in terms of Fredholm determinant [20] and not from a direct evaluation of their formula in Eqs. (161), (162), which we recall in Appendix D, in terms of an isomonodromic system associated to a Painlevé II transcendent. As explained in Ref. [27], this formula (161), (162), at variance with the one in terms of Fredholm determinant [27], can not be determined to arbitrary accuracy. This comparison shows a good agreement between the numerical evaluation of our formula in Eq. (108) and the values tabulated by WBF in [27]. In Fig. 5 we have also shown the data for $\tilde{p}_{\text{typ}}(\tilde{r})$ which we have computed numerically by sampling 2×10^5 independent Hermitian matrices of size 1000×1000 . The results of this numerical computation is shown in green filling in Fig. 5 and shows a nice agreement with both formula. Finally, in Fig. 5, we have also plotted the asymptotic behaviors (26) which we have obtained for small \tilde{r} for large \tilde{r} (black dashed curve). A more precise analysis of the small and large \tilde{r} of the data of WBF [27] shows actually a very good agreement with our precise asymptotics (26).

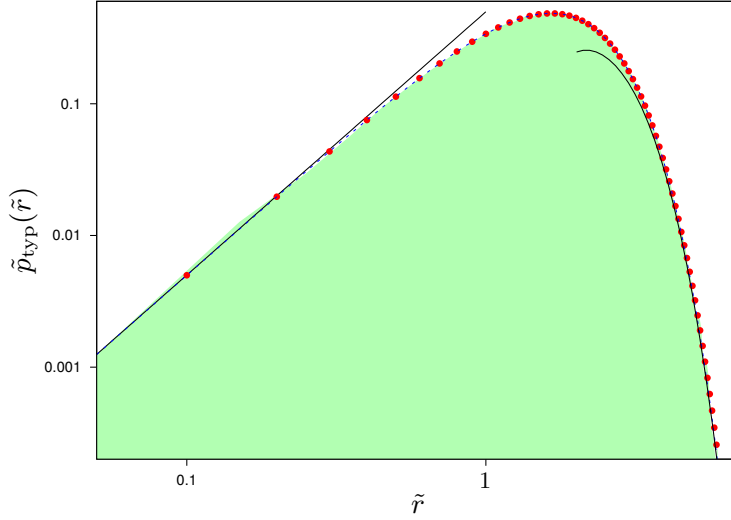


Fig. 5 Red dots: evaluation of our formula (108) for $\tilde{p}_{\text{typ}}(\tilde{r})$, black solid curve: WBF numerical result, green filling: numerical result with 2×10^5 realizations for a 1000×1000 GUE matrix, black dashed curve: asymptotic behaviors given in Eq. (26).

6 Conclusion and perspectives

To conclude, we have studied two different statistical characterizations of near-extreme eigenvalues of random matrices belonging to GUE: (i) the mean density of states $\rho_{\text{DOS}}(r, N)$ which is the mean density of eigenvalues located at a distance r from the largest one, λ_{max} and (ii) the PDF $p_{\text{GAP}}(r, N)$ of the first gap, i.e. the spacing between the two largest eigenvalues. The DOS has recently been studied in the context of i.i.d. random variables [4] and more recently for Brownian motion [28]. It is a natural object to characterize the important phenomenon of “crowding” near extremes. The study presented in this paper is a first attempt to characterize the “crowding” near the largest eigenvalue, λ_{max} , of random matrices. Here, we have performed a detailed study of the DOS for GUE and it would be interesting to address the same questions for Gaussian β -ensembles, for which we could only predict heuristically some asymptotic behaviors (121). It would also be interesting to study the density of states close to the smallest eigenvalue λ_{min} at the hard edge of Laguerre-Wishart matrices, where we would expect quite different behaviors.

We showed that these two quantities are actually related via the relation $p_{\text{GAP}}(r, N) = (N - 1)\rho_{\text{DOS}}(-r, N)$ and showed that both quantities can be conveniently expressed in terms of orthogonal polynomials which can be viewed as a generalization of Hermite polynomials defined on a semi-infinite interval (12). We showed, using a double scaling analysis, that these OP are related to the Lax pair of a Painlevé XXXIV equation, a result which can also be obtained using rather sophisticated method relying on Riemann-Hilbert techniques [38].

We could then characterize the PDF characterizing the typical fluctuations $\tilde{p}_{\text{typ}}(\tilde{r})$ of the first gap, which are of order $\mathcal{O}(N^{-1/6})$, in terms of the solution of this Lax pair. Using this expression, we could also derive precise asymptotics of this PDF (26). Our expression is different, and also simpler, than the one found previously by Witte, Bornemann and Forrester [27], and it would be very interesting to show explicitly that these two formulas do coincide. Note that similar relations arising in related contexts have recently emerged in [42, 47, 48]. It would also be natural to extend the present study of the first gap to other ensembles of RMT, and in particular to the Gaussian β -ensembles.

Finally, we have studied here the *typical* fluctuations of the first spacing at the edge, which are of order $\mathcal{O}(N^{-1/6})$. An interesting continuation of the present work would be, as it has been done in detail for the largest eigenvalue λ_{max} itself (for a recent review see [49]), to study the *atypically* large fluctuations of this first spacing, when they are of order $\mathcal{O}(N^{1/2})$. We leave these interesting open questions for future investigations.

Acknowledgements We would like to thank S. N. Majumdar for very stimulating discussions and a careful reading of our manuscript. We acknowledge support by ANR grant 2011-BS04-013-01 WALKMAT. G. S. also acknowledges support from Labex PALM (Project RANDMAT).

A Rewriting the Lax system in terms of rescaled variables

In this appendix, we derive two identities relating the Hastings-McLeod solution of Painlevé II with $\alpha = 0$, denoted by $q(x)$ (17), and the one with $\alpha = 1/2$, denoted by $q_{1/2}$ (62). We start with the relation between $q(x)$ and $q_{1/2}(x)$ [27, 45]

$$q_{1/2}(x) = -2^{-1/3} \frac{q'(-2^{-1/3}x)}{q(-2^{-1/3}x)}. \quad (123)$$

By taking the derivative with respect to (w.r.t.) x on both sides of Eq. (123), one obtains

$$q'_{1/2}(x) = 2^{-2/3} \frac{q''(-2^{-1/3}x)}{q(-2^{-1/3}x)} - 2^{-2/3} \left(\frac{q'(-2^{-1/3}x)}{q(-2^{-1/3}x)} \right)^2. \quad (124)$$

By combining Eq. (123) and Eq. (124), one obtains

$$x + 2q_{1/2}^2(x) + 2q'_{1/2}(x) = 2^{1/3} \left(2^{-1/3}x + \frac{q''(-2^{-1/3}x)}{q(-2^{-1/3}x)} \right). \quad (125)$$

Finally, using that $q(x)$ is solution of the Painlevé II equation (17) one obtains

$$q_{1/2}^2(x) + q'_{1/2}(x) + \frac{x}{2} = 2^{1/3} q^2(-2^{-1/3}x). \quad (126)$$

We now derive a second identity by considering the following function

$$J(x) = -x - 2q_{1/2}^2(x) + 2q'_{1/2}(x). \quad (127)$$

It is straightforward to check that $J(x)$ satisfies

$$J'(x) = -2q_{1/2}(x)J(x) - 2. \quad (128)$$

The solution of the homogenous equation, $J'(x) = -2q_{1/2}(x)J(x)$ is, using (123), $J(x) = A/q^2(-2^{-1/3}x)$. By varying the constant, one finds the solution of (128) under the form:

$$J(x) = \frac{A(x)}{q^2(-2^{-1/3}x)}, \quad A(x) = 2^{4/3} \int_{-2^{-1/3}x}^{\infty} q^2(u) du + a, \quad (129)$$

where a is a constant, independent of x . On the other hand, from the large x behavior $q_{1/2}(x) \sim 1/(2x)$, one sees that $J(x)$ in (127) behaves like $J(x) \sim -x$, when $x \rightarrow \infty$. This implies that the constant a in Eq. (129) is $a = 0$. Hence, one obtains a second identity:

$$-q_{1/2}^2(x) + q'_{1/2}(x) - \frac{x}{2} = -2^{1/3} \frac{\int_{-2^{-1/3}x}^{\infty} q^2(u) du}{q^2(-2^{-1/3}x)}. \quad (130)$$

One can then use these identities (126) and (130) to write the matrix elements of the matrices **A** and **B** in Eq. (68) in terms of $q(x)$ only – and not $q_{1/2}(x)$. This yields ultimately, with an appropriate change of variables, the expression of the matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ in Eq. (79).

B Expansion of the solution of the Lax system for small \tilde{r}

In this appendix, we give some details concerning the expansion of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ [which is actually similar to the one of $\tilde{\rho}_{\text{typ}}(\tilde{r})$, see Eq. (23)] for small argument \tilde{r} .

B.1 General structure of the psi-functions $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$ at small \tilde{r}

The small \tilde{r} expansion of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ necessitates the expansion of the solution of the Lax pair $\tilde{f}(\tilde{r}, x)$ and $\tilde{g}(\tilde{r}, x)$, which we recall are solutions of the system of differential equations

$$\frac{\partial}{\partial \tilde{r}} \begin{pmatrix} \tilde{f}(\tilde{r}, s) \\ \tilde{g}(\tilde{r}, s) \end{pmatrix} = \tilde{\mathbf{A}} \begin{pmatrix} \tilde{f}(\tilde{r}, s) \\ \tilde{g}(\tilde{r}, s) \end{pmatrix}, \quad \frac{\partial}{\partial s} \begin{pmatrix} \tilde{f}(\tilde{r}, s) \\ \tilde{g}(\tilde{r}, s) \end{pmatrix} = \tilde{\mathbf{B}} \begin{pmatrix} \tilde{f}(\tilde{r}, s) \\ \tilde{g}(\tilde{r}, s) \end{pmatrix}, \quad (131)$$

where $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are 2×2 matrices given by

$$\tilde{\mathbf{A}} = \begin{pmatrix} -\frac{q'(s)}{q(s)} & 1 + q^2(s)/\tilde{r} \\ -\tilde{r} - \frac{\int_s^\infty q^2(u) du}{q^2(s)} & \frac{q'(s)}{q(s)} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \frac{q'(s)}{q(s)} & -1 \\ \tilde{r} & -\frac{q'(s)}{q(s)} \end{pmatrix}, \quad (132)$$

where the solutions $\tilde{f}(r, s)$ and $\tilde{g}(r, s)$ are characterized by the asymptotic behaviors given in Eq. (80) in the text. We have already shown that $\tilde{f}(\tilde{r} = 0, x)$ exists, and in fact $\tilde{f}(\tilde{r} = 0, x) = 2^{-1/6} \sqrt{\pi} q(x)$ and given the \tilde{r} -dependence of the matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ one expects that $\tilde{f}(\tilde{r}, x)$ admits the following expansion

$$\tilde{f}(\tilde{r}, x) = 2^{-1/6} \sqrt{\pi} q(x) + \sum_{n=1}^{\infty} \tilde{r}^n \tilde{f}_n(x). \quad (133)$$

To obtain the small \tilde{r} expansion of $\tilde{g}(\tilde{r}, x)$, we show that it can be actually expressed in terms of $\tilde{f}(\tilde{r}, x)$. Using the relation (93) shown in the text, one obtains from (133) that $\tilde{g}(\tilde{r}, x)$ admits the following expansion

$$\tilde{g}(\tilde{r}, x) = \sum_{n=1}^{\infty} \tilde{r}^n \tilde{g}_n(x), \quad \tilde{g}_n(x) = -\frac{1}{q(x)} \int_x^\infty q(u) \tilde{f}_{n-1}(u) du, \quad n \geq 1. \quad (134)$$

By injecting the expansion of $\tilde{f}(\tilde{r}, x)$ (133) into the ' $\tilde{\mathbf{B}}$ -equation' satisfied by $\tilde{f}(\tilde{r}, x)$ [see Eqs. (131), (132)] one finds the following equations

$$\partial_x \tilde{f}_k(x) = \frac{q'(x)}{q(x)} \tilde{f}_k(x) - \tilde{g}_k(x), \quad (135)$$

where $\tilde{g}_k(x)$ defined in Eq. (134) can be expressed in terms of $\tilde{f}_{k-1}(x)$. Hence this set of equations (135) can be solved iteratively for successive values of k , starting from $k = 1$, to yield the first functions given in Eq. (100) in the text.

B.2 Lowest order expansion: calculation of the integral a_2

In this section of the appendix, we show that the amplitude a_2 defined through the rather complicated integral [see Eq. (103) in the text]

$$a_2 = \int_{-\infty}^{\infty} \left[(q' + qR)^2 - \frac{1}{4} (q^2 - R^2)^2 \right] \mathcal{F}_2 dx, \quad (136)$$

has actually a very simple expression, namely $a_2 = 1/2$. First we recall that

$$R(x) = \int_x^\infty q^2(u) du = \frac{\mathcal{F}_2'(x)}{\mathcal{F}_2(x)}. \quad (137)$$

This identity (137) is the crucial one as it allows us to compute this integral in (136), by using successive integration by parts. To this purpose, we first expand the squares in the integrand in (136) and decompose it as

$$a_2 = \int_{-\infty}^{\infty} \left(q'^2 - \frac{1}{4} q^4 \right) \mathcal{F}_2 dx + J_1 + J_2 + J_3, \quad (138a)$$

$$J_1 = 2 \int_{-\infty}^{\infty} q q' R \mathcal{F}_2 dx = 2 \int_{-\infty}^{\infty} q q' \mathcal{F}_2' dx, \quad (138b)$$

$$J_2 = \frac{3}{2} \int_{-\infty}^{\infty} q^2 R^2 \mathcal{F}_2 dx = \frac{3}{2} \int_{-\infty}^{\infty} q^2 R \mathcal{F}_2' dx, \quad (138c)$$

$$J_3 = -\frac{1}{4} \int_{-\infty}^{\infty} R^4 \mathcal{F}_2 dx = -\frac{1}{4} \int_{-\infty}^{\infty} R^3 \mathcal{F}_2' dx, \quad (138d)$$

where we used the shorthand notations $q \equiv q(x)$, $R \equiv R(x)$ and $\mathcal{F}_2 \equiv \mathcal{F}_2(x)$. We compute the integral J_1 in (138b) by using an integration by part [integrating $\mathcal{F}_2'(x)$], which yields

$$J_1 = -2 \int_{-\infty}^{\infty} (q'^2 + 2q^4 + xq^2) \mathcal{F}_2 dx, \quad (139)$$

where we have used that $q(x)$ is solution of the Painlevé II equation (17) as well as the asymptotic behavior of $\mathcal{F}_2(x)$ for large negative argument (116). Similarly, the integral J_2 in (138c) can be transformed by using a similar integration by part [again integrating $\mathcal{F}_2'(x)$]. This yields

$$J_2 = \frac{3}{2} \int_{-\infty}^{\infty} q^4 \mathcal{F}_2 dx - \frac{3}{2} J_1 = \int_{-\infty}^{\infty} \left(3q'^2 + \frac{15}{2} q^4 + 3xq^2 \right) \mathcal{F}_2 dx. \quad (140)$$

The integral J_3 in (138d) can be transformed using the same procedure as

$$J_3 = -\frac{J_2}{2} = -\frac{1}{2} \int_{-\infty}^{\infty} \left(3q'^2 + \frac{15}{2} q^4 + 3xq^2 \right) \mathcal{F}_2 dx. \quad (141)$$

Combining Eqs. (138a)- (141) one obtains

$$a_2 = \frac{1}{2} \int_{-\infty}^{\infty} (q'^2 - q^4 - xq^2) \mathcal{F}_2 dx. \quad (142)$$

Finally, this last integral can be computed exactly by using the identity:

$$R(x) = \int_x^{\infty} q^2(u) du = q'^2 - q^4 - xq^2, \quad (143)$$

which can be checked easily by taking derivative with respect to x on both sides, and using that $q(x)$ is solution of the Painlevé II equation (17). Hence using this identity (143), a_2 in (142) can be computed as

$$a_2 = \frac{1}{2} \int_{-\infty}^{\infty} R(x) \mathcal{F}_2(x) dx = \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{F}_2'(x) dx = \frac{1}{2}, \quad (144)$$

as given in Eq. (103) in the text.

C Large \tilde{r} expansion of $\tilde{p}_{\text{typ}}(\tilde{r})$: beyond the leading order

In this appendix, we analyze in detail the large \tilde{r} asymptotic of $\tilde{p}_{\text{typ}}(\tilde{r})$. We obtain in particular the first sub-leading corrections to the leading term obtained in the text, in Eq. (118), yielding the rather precise asymptotics for $\tilde{p}_{\text{typ}}(\tilde{r})$ given in Eq. (119).

This expansion, beyond leading order, requires the determination of the first correction, of order $\mathcal{O}(\tilde{r}^{-1/2})$ to the asymptotic behavior of $\tilde{f}(-\tilde{r}, x)$, for large \tilde{r} , given in Eq. (110), which we first focus on. We expect, from (110), the following asymptotic behavior valid for large \tilde{r} :

$$\tilde{f}(-\tilde{r}, x) = \frac{1}{2^{7/6}} \tilde{r}^{-1/4} e^{-\frac{2}{3} \tilde{r}^{2/3} - x\sqrt{\tilde{r}}} \left(1 + \frac{1}{\sqrt{\tilde{r}}} F_1(x) + o(\tilde{r}^{-1/2}) \right). \quad (145)$$

To compute the function $F_1(x)$, we use that $\tilde{f}(-\tilde{r}, x)$ satisfies the following Schrödinger equation

$$\partial_x^2 \tilde{f}(-\tilde{r}, x) - (x + 2q^2(x)) \tilde{f}(-\tilde{r}, x) = \tilde{r} \tilde{f}(-\tilde{r}, x). \quad (146)$$

By injecting the asymptotic expansion (145) into Eq. (146), one obtains $F_1(x)$ as

$$F_1(x) = -\frac{1}{2} \int_{-\infty}^x [u + 2q^2(u)] du, \quad (147)$$

where we have used that $\lim_{x \rightarrow -\infty} F_1(x) = 0$. Although this is a reasonable assumption we have not been able to establish it rigorously. In the following, we will need the behavior of $F_1(x)$ for large negative argument. It can be obtained from the large negative argument of $q(s)$:

$$q(s) = \sqrt{-\frac{s}{2}} \left(1 + \frac{1}{8s^3} + \mathcal{O}(s^{-6}) \right), \text{ when } s \rightarrow -\infty, \quad (148)$$

which yields, for $F_1(x)$ in (147)

$$F_1(x) = -\frac{1}{8x} + o(|x|^{-1}), \text{ when } x \rightarrow -\infty. \quad (149)$$

More generally, the analysis of Eq. (146) suggests that $\tilde{f}(-\tilde{r}, x)$ admits an expansion of the form

$$\tilde{f}(-\tilde{r}, x) = \frac{1}{2^{7/6}} \tilde{r}^{-1/4} e^{-\frac{2}{3}\tilde{r}^{2/3} - x\sqrt{\tilde{r}}} \left(1 + \sum_{n=1}^{\infty} \frac{1}{\tilde{r}^{n/2}} F_n(x) \right), \quad F_n(x) \underset{x \rightarrow -\infty}{\sim} \alpha_n |x|^{-n}. \quad (150)$$

Equipped with this asymptotic expansion (150) we can now compute the large \tilde{r} asymptotics of $\tilde{p}_{\text{typ}}(\tilde{r})$ beyond leading order. As this was done in the main text, the starting point of our analysis is the following expression

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \frac{1}{r^2} q^2(x) \tilde{g}^2(-\tilde{r}, x) \right] \mathcal{F}_2(x) dx. \quad (151)$$

We analyze the first term in (151) – the integral involving $\tilde{f}^2(-\tilde{r}, x)$ – by injecting the large \tilde{r} expansion obtained above (150). It yields

$$\frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \tilde{f}^2(-\tilde{r}, x) = \frac{1}{4\pi} \frac{1}{\tilde{r}^{1/2}} e^{-\frac{4}{3}\tilde{r}^{3/2}} \int_{-\infty}^{\infty} e^{-2x\sqrt{\tilde{r}}} \left(1 + \frac{2}{\sqrt{r}} F_1(x) + o(\tilde{r}^{-1/2}) \right) \mathcal{F}_2(x) dx. \quad (152)$$

As we have seen before [see Eq. (117) in the text], the integral over x in (153) is dominated by the region where $x < 0$ where, as shown in the main text, one can use the saddle point method. Indeed, as shown previously in (117), the saddle point is reached for $x^* = 2\sqrt{2}\tilde{r}^{1/4}$. It is thus natural to perform the change of variable $x = u\tilde{r}^{1/4}$ and use the asymptotic behavior of $F_1(x)$ for large negative argument given in (149) as well as the one for $\mathcal{F}_2(x)$ [46]:

$$\mathcal{F}_2(x) = \tau_2 |x|^{-1/8} e^{-\frac{1}{12}|x|^3} \left(1 + \frac{3}{2^6|x|^3} + \mathcal{O}(|x|^{-6}) \right), \quad \tau_2 = 2^{1/24} e^{\zeta'(-1)}, \quad (153)$$

to obtain after straightforward (though tedious) manipulations the following expansion:

$$\frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \tilde{f}^2(-\tilde{r}, x) = A \exp \left(-\frac{4}{3}\tilde{r}^{3/2} + \frac{8}{3}\sqrt{2}\tilde{r}^{3/4} \right) \tilde{r}^{-21/32} \left(1 + \frac{131\sqrt{2}}{1536}\tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2}) \right), \quad (154)$$

with $A = 2^{-91/48} e^{\zeta'(-1)} / \sqrt{\pi}$.

We now analyze the second term in Eq. (151), which involves $\tilde{g}(-\tilde{r}, x)$ that does not contribute to leading order when $\tilde{r} \rightarrow \infty$. To analyze this term, it is sufficient to expand $\tilde{g}(-\tilde{r}, x)$ using Eq. (110) as well as $\mathcal{F}_2(x)$ using Eq. (153) to leading order. One can then perform a large \tilde{r} expansion of this term using again the saddle point method, as shown in the main text (117). One obtains, after some manipulations:

$$-\frac{2^{1/3}}{\pi \tilde{r}^2} \int_{-\infty}^{\infty} \left[q^2(x) \tilde{g}^2(-\tilde{r}, x) \right] \mathcal{F}_2(x) dx = A \exp \left(-\frac{4}{3}\tilde{r}^{3/2} + \frac{8}{3}\sqrt{2}\tilde{r}^{3/4} \right) \tilde{r}^{-21/32} \times \left(-\sqrt{2}\tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2}) \right) \quad (155)$$

Finally, combining these asymptotic expansions (154) and (155) one obtains from (151)

$$\tilde{p}_{\text{typ}}(\tilde{r}) = A \exp \left(-\frac{4}{3}\tilde{r}^{3/2} + \frac{8}{3}\sqrt{2}\tilde{r}^{3/4} \right) \tilde{r}^{-21/32} \left(1 - \frac{1405\sqrt{2}}{1536}\tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2}) \right), \quad (156)$$

as announced in the text in Eq. (119).

D The formula for the first gap, $\tilde{p}_{\text{typ}}(\tilde{r})$, obtained by Witte, Bornemann and Forrester (WBF)

In Ref. [27], WBF obtained a formula for $\tilde{p}_{\text{typ}}(\tilde{r})$ in terms of the components of a solution of a particular isomonodromic problem relating to the Painlevé II equation. In this paper, they used a method which differs significantly for the one developed in the present paper. The starting point of their computation [27] is a formula, obtained in Ref. [50], for the joint PDF of the first and second smallest eigenvalues at the hard edge of unitary ensembles. They could then use a transformation which relates the hard edge to the soft edge [51] to obtain an expression for $\tilde{p}_{\text{typ}}(\tilde{r})$ in terms of the Hamiltonian system for Painlevé II, denoted by $\{q, p; t, H\}$. We

introduce the parameters α and α_1 satisfying the relations $\alpha = \alpha_1 - 1/2$. In the present case, $\alpha_1 = 2$, implying $\alpha = 3/2$. The Painlevé II Hamiltonian is

$$H = -\frac{1}{2} \left(2q_\alpha^2 - p_\alpha + t \right) p_\alpha - \alpha_1 q_\alpha, \quad (157)$$

while the corresponding equations of motion are

$$\partial_t q_\alpha = \partial_{p_\alpha} H = p_\alpha - q_\alpha^2 - \frac{1}{2}t, \quad \partial_t p_\alpha = \partial_{q_\alpha} H = 2q_\alpha p_\alpha + \alpha_1, \quad (158)$$

from which $q_\alpha(t)$ satisfies the Painlevé II equation with parameter α

$$\partial_t^2 q_\alpha = 2q_\alpha^3 + tq_\alpha + \alpha. \quad (159)$$

Note that H satisfies the second-order second degree differential equation of Jimbo-Miwa-Okamoto σ form for PII with $\alpha_1 = 2$, $\alpha = \frac{3}{2}$:

$$\ddot{H}^2 + 4\dot{H}^3 + 2\dot{H}[t\dot{H} - H] = 1. \quad (160)$$

In terms of the Hamiltonian system (157) (with $\alpha_1 = 2$), the expression of $\tilde{p}_{\text{typ}}(\tilde{r})$ obtained in Ref. [27] reads

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \int_{-\infty}^{\infty} p_{(2)}^{\text{soft}}(t, t - \tilde{r}) dt, \quad (161)$$

with

$$p_{(2)}^{\text{soft}}(t, t - \tilde{r}) = \frac{t^{-5/2}}{4\pi} p_{(1)}^{\text{soft}}(t) \exp\left(-\frac{4}{3}t^{3/2}\right) \exp\left(\int_{2^{1/3}t}^{\infty} dy \left\{ (2q_{3/2} + \frac{4}{p_{3/2}})(-y) - \sqrt{2y} - \frac{5}{2y} \right\}\right) \\ \times (U\partial_x V - V\partial_x U)(-2^{1/3}\tilde{r}; -2^{1/3}t). \quad (162)$$

In Eq. (162), $p_{(1)}^{\text{soft}}(t)$ is given by [27]

$$p_{(1)}^{\text{soft}}(t) = K^{\text{soft}}(t, t) \exp\left(-\int_s^{\infty} dt \left(\sigma_{\text{II}}(t) - \frac{d}{dt} \log K^{\text{soft}}(t, t)\right)\right), \quad (163)$$

where $K^{\text{soft}}(x, y)$ is the Airy kernel

$$K^{\text{soft}}(x, y) = \frac{\text{Ai}(x)\text{Ai}'(y) - \text{Ai}(y)\text{Ai}'(x)}{x - y} \quad (164)$$

and

$$\sigma_{\text{II}}(t) = -2^{1/3}H(-2^3t), \quad (165)$$

where the Hamiltonian H is given in Eq. (157). Finally, in Eq. (162), $U \equiv U(x, t)$ and $V = V(x, t)$ are solutions of the following Lax pair:

$$\partial_x \begin{pmatrix} U \\ V \end{pmatrix} = \left\{ \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & 0 \end{pmatrix} x + \begin{pmatrix} -q_{3/2} - \frac{2}{p_{3/2}} & -1 \\ \frac{1}{2}(t - p_{3/2}) + [q_{3/2} + \frac{2}{p_{3/2}}]^2 & q_{3/2} + \frac{2}{p_{3/2}} \end{pmatrix} + \begin{pmatrix} 1 & p_{3/2} \\ 0 & -1 \end{pmatrix} \frac{1}{x} \right\} \begin{pmatrix} U \\ V \end{pmatrix}, \quad (166)$$

$$\partial_t \begin{pmatrix} U \\ V \end{pmatrix} = \left\{ \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{pmatrix} x + \begin{pmatrix} 0 & 1 \\ 0 & -2[q_{3/2} + \frac{2}{p_{3/2}}] \end{pmatrix} \right\} \begin{pmatrix} U \\ V \end{pmatrix}. \quad (167)$$

$$(168)$$

which satisfy the boundary conditions for $t, x, t - x \rightarrow -\infty$

$$U(x, t) \sim -t \frac{\text{Ai}(2^{-1/3}(x - t))}{\text{Ai}(-2^{-1/3}t)} \quad (169)$$

$$V(x, t) \sim t \frac{2^{-1/3}\text{Ai}'(2^{-1/3}(x - t)) + \sqrt{-\frac{t}{2}}\text{Ai}(2^{-1/3}(x - t))}{\text{Ai}(-2^{-1/3}t)}. \quad (170)$$

References

1. J.-P. Bouchaud, M. Mézard, *Universality Classes for Extreme Value Statistics*, J. Phys. A **30**, 7997 (1997).
2. D. S. Dean, S. N. Majumdar, *Extreme Value Statistics of Hierarchically Correlated Variables: Deviation from Gumbel Statistics and Anomalous Persistence*, Phys. Rev. E **64**, 046121 (2001).
3. P. Le Doussal, C. Monthus, *Exact solutions for the statistics of extrema of some random 1D landscapes, Application to the equilibrium and the dynamics of the toy model*, Physica A **317**, 140 (2003).
4. S. Sabhapandit, S. N. Majumdar, *Density of near-extreme events*, Phys. Rev. Lett. **98**(14), 140201 (2007).
5. D. S. Fisher, D. Huse, *Nonequilibrium dynamics of spin glasses*, Phys. Rev. B **38**, 373 (1988).
6. D. S. Fisher, D. Huse, *Equilibrium behavior of the spin-glass ordered phase*, Phys. Rev. B **38**, 386 (1988).
7. C. Monthus, P. Le Doussal, *Low-temperature properties of some disordered systems from the statistical properties of nearly degenerate two-level excitations*, Eur. Phys. J. B **41**, 535 (2004).
8. C. Monthus, T. Garel, *Typical versus averaged overlap distribution in Spin-Glasses: Evidence for the droplet scaling theory*, Phys. Rev. B **88**, 134204 (2013).
9. F. Omori, *On the aftershocks of earthquakes*, J. Coll. Sci. Imp. Univ. Tokyo **7**, 111 (1894).
10. D. Vere-Jones, *A note on the statistical interpretation of Bath's Law*, Bull. Seism. Soc. Am. **59**, 1535 (1969).
11. A. M. Petersen, F. Wang, S. Havlin, H. E. Stanley, *Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws*, Phys. Rev. E **82**, 036114 (2010).
12. S. Sabhapandit, S. N. Majumdar, S. Redner, *Crowding at the Front of the Marathon Packs*, J. Stat. Mech. L03001, (2008).
13. H. A. David, H. N. Nagaraja, *Order Statistics* (third ed.), Wiley, New Jersey (2003).
14. E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, (1958).
15. E. Brunet, B. Derrida, *Statistics at the tip of a branching random walk and the delay of traveling waves*, Europhys. Lett. **87**, 60010 (2009).
16. E. Brunet, B. Derrida, *A branching random walk seen from the tip*, J. Stat. Phys. **143**, 420 (2011).
17. N. R. Moloney, K. Ozogány, Z. Rácz, *Order statistics of $1/f^\alpha$ signals*, Phys. Rev. E **84**, 061101 (2011).
18. G. Schehr, S. N. Majumdar, *Universal order statistics of random walks*, Phys. Rev. Lett. **108**, 040601 (2012).
19. S. N. Majumdar, P. Mounaix, G. Schehr, *Exact Statistics of the Gap and Time Interval between the First Two Maxima of Random Walks and Lévy Flights*, Phys. Rev. Lett. **111**, 070601 (2013).
20. C. A. Tracy, H. Widom, *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys. **159**, 151 (1994).
21. C. A. Tracy, H. Widom, *On the orthogonal and symplectic ensembles*, Commun. Math. Phys. **177**, 727 (1996).
22. M. Dieng and C. A. Tracy, *Application of Random Matrix Theory to Multivariate Statistics in Random matrices, random processes and Integrable systems*, Ed J. Harnad, Springer, NY 2011; M. Dieng preprint arxiv:0506586.
23. S. Gustavsson, *Gaussian fluctuations of eigenvalues in the GUE*, Ann. Inst. H. Poincaré, Probab. Statist. **41**, 151 (2005).
24. S. O'Rourke, *Gaussian Fluctuations of Eigenvalues in Wigner Random Matrices*, J. Stat. Phys. **138**, 1045 (2010).
25. M. L. Mehta, *Random Matrices*, 2nd Edition, Academic Press (1991).
26. P. J. Forrester, *Log-gases and random matrices*, Princeton University Press, Princeton, NJ, (2010).
27. N. S. Witte, F. Bornemann, P. J. Forrester, *Joint distribution of the first and second eigenvalues at the soft edge of unitary ensembles*, Nonlinearity **26**, 1799 (2013).
28. A. Perret, A. Comtet, S. N. Majumdar, G. Schehr, *Near-extreme statistics of Brownian motion*, preprint arXiv:13056490, to be published in Phys. Rev. Lett.
29. A. G. Pakes, F. W. Steutel, *On the number of records near the maximum*, Aust. J. Stat. **39**, 179 (1997).
30. A. G. Pakes, Y. Li, *Limit laws for the number of near maxima via the Poisson approximation*, Statist. Probab. Lett. **40**, 395 (1998).
31. P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nucl. Phys. B **402**(3), 709 (1993).
32. M. Bowick, E. Brézin, *Universal scaling of the tail of the density of eigenvalues in random matrix models*, Phys. Lett. B **268**, 21 (1991).
33. S. N. Majumdar, G. Schehr, D. Villamaina, *Large deviations of the top eigenvalue of large Cauchy random matrices*, P. Vivo, J. Phys. A: Math. Theor. **46** 022001 (2013).
34. C. Nadal, S. N. Majumdar, *A simple derivation of the Tracy-Widom distribution of the maximal eigenvalue of a Gaussian unitary random matrix*, J. Stat. Mech., P04001 (2011).
35. G. Szegő, *Orthogonal polynomials*, American Mathematical Society, Providence R.I., 4th edition, (1975).
36. G. Akemann, M. Atkin, *Higher Order Analogues of Tracy-Widom distributions via the Lax method* J. Phys. A: Math. Theor. **46**, 015202 (2013).
37. M. Atkin, S. Zohren, *Instantons and extreme value statistics of random matrices*, preprint arXiv: 1307.3118.
38. T. Claeys, A. B. J. Kuijlaars, *Universality in unitary random matrix ensembles when the soft edge meets the hard edge*, in *Integrable systems and random matrices*, pp. 265–279, Contemp. Math., 458, Amer. Math. Soc., Providence, RI, (2008).
39. A. Borodin, A. Soshnikov, *Janossy densities I. Determinantal ensembles*, J. Stat. Phys. **113**, 595 (2003).
40. V. Periwal, D. Shevitz, *Unitary-matrix models as exactly solvable string theories*, Phys. Rev. Lett. **64**, 1326 (1990).
41. D. J. Gross, A. Matytsin, *Instanton induced large N phase transitions in two and four dimensional QCD*, Nucl. Phys. B **429**, 50 (1974).

-
42. G. Schehr, *Extremes of N vicious walkers for large N : application to the directed polymer and KPZ interfaces*, J. Stat. Phys. **149**, 385 (2012).
 43. A. S. Fokas, A. R. Its, A. A. Kapaev, V. Yu. Novokshenov, *Painlevé transcendents: the Riemann-Hilbert Approach*, Mathematical Surveys and Monographs Vol. 128, Amer. Math. Soc. Providence, R. I. (2006).
 44. This was done using the tabulation of $q(x)$, $R(x)$ and $\mathcal{F}_2(x)$ which can be found on the webpage of M. Prähofer: <http://www-m5.ma.tum.de/KPZ>.
 45. P. A. Clarkson, N. Joshi, A. Pickering, *Bäcklund transformations for the second Painlevé hierarchy: a modified truncation approach*, Inverse Probl. **15**, 175 (1999).
 46. J. Baik, R. Buckingham, J. DiFranco, *Asymptotics of Tracy-Widom distributions and the total integral of a Painlevé II function*, Comm. Math. Phys. **280**, 463 (2008).
 47. J. Baik, K. Liechty, G. Schehr, *On the joint distribution of the maximum and its position of the Airy₂ process minus a parabola*, J. Math. Phys. **53**, 083303 (2012).
 48. S. Delvaux, *The tacnode kernel: equality of Riemann-Hilbert and Airy resolvent formulas*, preprint arXiv:1211.4845.
 49. S. N. Majumdar, G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, preprint arXiv:1311.0580.
 50. P. J. Forrester, N. S. Witte, *The distribution of the first eigenvalues spacing at the hard edge of the Laguerre unitary ensemble*, Kyushu J. Math. **61**, 457 (2007).
 51. A. Borodin, P. J. Forrester, *Increasing subsequences and the hard-to-soft edge transition in matrix ensembles*, J. Phys. A **36**, 2963 (2003).

The density of eigenvalues seen from the soft edge of random matrices in the Gaussian β -ensembles.

[127] A. Perret et G. Schehr (2015).

Prépublication arXiv :1506.00245

Pour Random Matrix Theory : Foundations and Applications, Kraków

The density of eigenvalues seen from the soft edge of random matrices in the Gaussian β -ensembles *

ANTHONY PERRET, GRÉGORIE SCHEHR

Univ. Paris-Sud – Paris 11, CNRS, Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS), 91405 Orsay Cedex, France

(Received May 28, 2015)

We characterize the phenomenon of “crowding” near the largest eigenvalue λ_{\max} of random $N \times N$ matrices belonging to the Gaussian β -ensemble of random matrix theory, including in particular the Gaussian orthogonal ($\beta = 1$), unitary ($\beta = 2$) and symplectic ($\beta = 4$) ensembles. We focus on two distinct quantities: (i) the density of states (DOS) near λ_{\max} , $\rho_{\text{DOS}}(r, N)$, which is the average density of eigenvalues located at a distance r from λ_{\max} (or the density of eigenvalues seen from λ_{\max}) and (ii) the probability density function of the gap between the first two largest eigenvalues, $p_{\text{GAP}}(r, N)$. Using heuristic arguments as well as numerical simulations, we generalize our recent exact analytical study of the Hermitian case (corresponding to $\beta = 2$). We also discuss some applications of these two quantities to statistical physics models.

PACS numbers: 02.10.Yn, 05.40.-a

1. Introduction

During the last 20 years, there has been an important activity, both in mathematics and in physics, aiming at describing the fluctuations of the largest eigenvalue in ensembles of random matrices [1, 2, 3, 4, 5, 6, 7]. The most studied ones in this context, which we also focus on here, are probably the Gaussian β -ensembles where the joint probability density function (PDF) of the N real eigenvalues $\lambda_1, \dots, \lambda_N$ is given by:

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j} |\lambda_i - \lambda_j|^\beta \exp \left(-\frac{\beta}{2} \sum_{i=1}^N \lambda_i^2 \right), \quad (1)$$

where the normalization constant is $Z_N = (2\pi)^{N/2} \beta^{-N/2 - \beta N(N-1)/4} \Gamma(1 + \beta/2)^N \prod_{j=1}^N \Gamma(1 + \beta j/2)$ and where $\beta > 0$ is the Dyson index that can take

* for Random Matrix Theory : Foundations and Applications, Kraków.

any real value. The classical values correspond to $\beta = 1, 2$ and 4 , associated respectively to the Gaussian Orthogonal Ensemble (GOE), the Gaussian Unitary Ensemble (GUE) and the Gaussian Symplectic Ensemble (GSE). Note that for arbitrary β , it is possible to associate a matrix model to (1) (namely tridiagonal random matrices introduced in [8]). The fluctuations of the largest eigenvalues $\lambda_{\max} = \max_{1 \leq i \leq N} \lambda_i$, characterized by its cumulative distribution $F_N(y) = \text{Proba.}[\lambda_{\max} \leq y]$, are now well understood. Indeed, we have now a precise characterization of the typical fluctuations of λ_{\max} , when $|\lambda_{\max} - \sqrt{2N}|$ is of order $\mathcal{O}(N^{-1/6})$, which are described the TW distributions [1, 2] as well as of the large deviations of λ_{\max} , when $|\lambda_{\max} - \sqrt{2N}|$ is of order $\mathcal{O}(\sqrt{N})$, where $F_N(y)$ is described by large deviations functions [3, 4, 5, 6, 7] (both left and right tails).

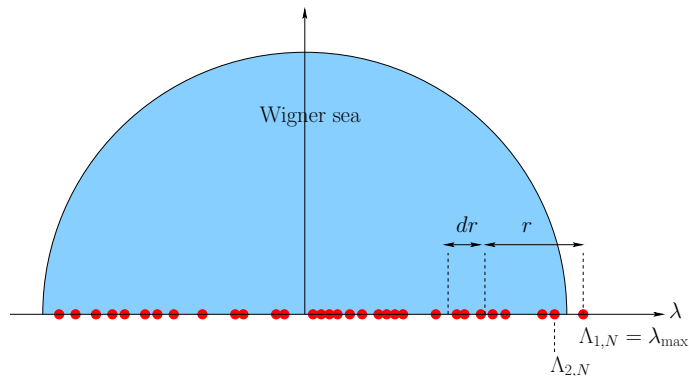


Figure 1. Different quantities characterizing the "crowding" near the largest eigenvalue λ_{\max} studied in this paper: (i) the mean density of states $\rho_{\text{DOS}}(r, N)$ such that $\rho_{\text{DOS}}(r, N)dr$ is the mean number of eigenvalues located in the interval $[\lambda_{\max} - r - dr, \lambda_{\max} - r]$ and (ii) the PDF $p_{\text{GAP}}(r, N)$ of the spacing between the two largest eigenvalues, $p_{\text{GAP}}(r, N)dr = \text{Pr.}[(\Lambda_{1,N} - \Lambda_{2,N}) \in [r, r + dr]]$.

However in several situations, some observables related to such spectral statistics (1), might be sensitive not only to λ_{\max} but also to near-extreme eigenvalues, whose amplitudes are smaller but close to this largest eigenvalue. This general feature, not restricted to random matrices, has led physicists to study the phenomenon of "crowding" close to the maximum [9]. This was studied in detail for the case of independent and identical random variables [9] and more recently for Brownian motion [10]. In Ref. [11] we proposed to characterize this phenomenon of "crowding" in random matrix models, by studying the density of near extreme eigenvalues, also called the density of

states, $\rho_{\text{DOS}}(r, N)$, defined as [9, 11] (see Fig. 1)

$$\rho_{\text{DOS}}(r, N) = \frac{1}{N-1} \sum_{\substack{i=1 \\ i \neq i_{\text{max}}}}^N \langle \delta(\lambda_{\text{max}} - \lambda_i - r) \rangle, \quad (2)$$

where i_{max} is such that $\lambda_{i_{\text{max}}} = \lambda_{\text{max}}$ and $\langle \cdots \rangle$ means an average taken with the weight in (1). It is normalized according to

$$\int_0^\infty dr \rho_{\text{DOS}}(r, N) = 1. \quad (3)$$

Incidentally, we showed that the DOS is related to another interesting observable characterizing also the crowding to λ_{max} , namely the PDF of the gap between the two largest eigenvalues [11]. Let us denote by $\lambda_{\text{max}} = \Lambda_{1,N} \geq \Lambda_{2,N} \geq \cdots \geq \Lambda_{N,N}$ and by $d_{1,N} = \Lambda_{1,N} - \Lambda_{2,N}$ the first gap (see Fig. 1). Its PDF is denoted by $p_{\text{GAP}}(r, N)$, such that $\text{Pr.}[d_{1,N} \in [r, r + dr]] = p_{\text{GAP}}(r, N)dr$. It is then possible to show the following relation:

$$p_{\text{GAP}}(r, N) = (N-1)\rho_{\text{DOS}}(-r, N), \quad (4)$$

and we refer the reader to Ref. [11] for the derivation of this relation.

In Ref. [11], we focused essentially on the case of GUE ($\beta = 2$) where we computed exactly these two quantities $\rho_{\text{DOS}}(r, N)$ and $p_{\text{GAP}}(r, N)$ both for finite N and in the large N limit, using a method based on (semi-classical) orthogonal polynomials, which were previously introduced in Ref. [12]. Here, we generalize these results to the case of generic β , which we mainly study, for large N , using heuristic arguments, that are confirmed by numerical simulations (for $\beta = 1, 2$ and $\beta = 4$). Before presenting our results we discuss two applications of these two near extreme observables (2, 4).

2. Two applications of $\rho_{\text{DOS}}(r, N)$ and $p_{\text{GAP}}(r, N)$

As we discuss it here, $\rho_{\text{DOS}}(r, N)$ and $p_{\text{GAP}}(r, N)$ play an important role in the minimization of quadratic forms on the sphere [13] and in the study of a spherical mean-field spin-glass [14] respectively.

2.1. Minimizing a quadratic form on the N -dimensional sphere

Let us consider the problem which consists in minimizing a random quadratic form on the N -dimensional sphere S_N :

$$H[\vec{s}] = -\frac{1}{2} \sum_{i,j=1}^N J_{i,j} s_i s_j, \quad \vec{s}^2 = \sum_{i=1}^N s_i^2 = N, \quad (5)$$

where \mathbf{J} is a matrix belonging to the Gaussian β -ensemble (with $\beta = 1, 2$ being the more natural values in this case) – and here we choose the matrix elements $J_{i,j}$ of zero mean and variance of order $\mathcal{O}(1/N)$ ¹. To take into account the spherical constraint (5) we introduce a Lagrange multiplier z such that we have to minimize

$$\tilde{H}[\vec{s}, z] = -\frac{1}{2} \sum_{i,j=1}^N J_{i,j} s_i s_j + z (\vec{s}^2 - N) \quad (6)$$

with respect to \vec{s} and z . It is then straightforward to show that

$$\min_{\vec{s}, z} \tilde{H}[\vec{s}, z] = \tilde{H}[\vec{s}_{\max}, z_{\max}] = -N \frac{\lambda_{\max}}{2}, \quad (7)$$

where \vec{s}_{\max} and z_{\max} are such that

$$\begin{cases} \mathbf{J} \vec{s}_{\max} = \lambda_{\max} \vec{s}_{\max}, \\ z_{\max} = \frac{\lambda_{\max}}{2}. \end{cases} \quad (8)$$

If we now look at the Hessian matrix $\delta^2 \tilde{H} / \delta s_i \delta s_j$ evaluated at the minimum \vec{s}_{\max}, z_{\max} , one can show that its spectrum is given by

$$\text{Sp} \left[\frac{\delta^2 \tilde{H}}{\delta s_i \delta s_j} \Big|_{\vec{s}_{\max}, z_{\max}} \right] = \{0, \lambda_{\max} - \lambda_1, \lambda_{\max} - \lambda_2, \dots, \lambda_{\max} - \lambda_N\}. \quad (9)$$

Hence the DOS, $\rho_{\text{DOS}}(r, N)$ in Eq. (2) is the average density of eigenvalues of the Hessian evaluated at the minimum (except the trivial eigenvalue 0). It is thus natural to expect that the DOS plays an important role in the relaxational dynamics of a system driven by such a quadratic form (5), which corresponds precisely to the fully connected p -spin spherical spin-glass model with $p = 2$ (where the s_i 's correspond to spin variables coupled to each other via the matrix \mathbf{J}) [15, 16, 17].

2.2. Overlap distribution in the fully connected $p = 2$ -spherical spin glass model

In Ref. [14], the authors studied the equilibrium properties of the $p = 2$ -spherical spin glass model (or spherical Sherrington-Kirkpatrick model) described by the partition function associated to the above Hamiltonian (5)

$$Z = \int_{-\infty}^{\infty} ds_1 \cdots \int_{-\infty}^{\infty} ds_N e^{-\frac{1}{2T} \sum_{i,j} J_{i,j} s_i s_j} \delta \left(\sum_{i=1}^N s_i^2 - N \right), \quad (10)$$

¹ note that the joint PDF (1) corresponds instead to matrix elements of order $\mathcal{O}(1)$.

where the δ function ensures the spherical constraint (5). They focused on the overlap Q , which is the order parameter characterizing the spin-glass order, defined by

$$Q = \sum_{i=1}^N s_i^{(1)} s_i^{(2)}, \quad (11)$$

where $s_i^{(1)}$ and $s_i^{(2)}$ represent the spins at site i in two distinct equilibrium configurations with the same realization of the couplings $J_{i,j}$. In Ref. [14] it was shown that, for fixed couplings $J_{i,j}$, the distribution of Q (with respect to thermal fluctuations) is related, at low temperature T , to the first gap between the two largest eigenvalues $\Lambda_{1,N}$ and $\Lambda_{2,N}$ of the matrix \mathbf{J}

$$P(Q, N) \sim \frac{1}{N} \tilde{P}\left(q = \frac{Q}{N}\right), \quad \tilde{P}(q) = e^{-\frac{N}{2T}(\Lambda_{1,N} - \Lambda_{2,N})(1 - q^2)}. \quad (12)$$

Hence we see from (12) that the full distribution of the overlap in this model is directly related to the PDF of the first gap $d_{1,N} = \Lambda_{1,N} - \Lambda_{2,N}$ of Gaussian random matrices $J_{i,j}$, the natural ensemble being here GOE ($\beta = 1$). Indeed, after averaging the distribution $P(Q, N)$ in (12) over the random couplings $J_{i,j}$, one obtains nothing else but the Laplace transform (with Laplace parameter $\propto (1 - q^2)/T$) of the PDF of the first gap, $p_{\text{GAP}}(r, N)$, studied here [see Eq. (4)].

3. Two different scaling regimes: bulk and edge

Computing the DOS $\rho_{\text{DOS}}(r, N)$ for finite N is, for a generic value of β in (1), a very challenging task. This could be done for the special case $\beta = 2$ in Ref. [11] using the method of orthogonal polynomials, which unfortunately can not be extended to other values of β . In spite of this difficulty, the main features of $\rho_{\text{DOS}}(r, N)$, for large N , can be characterized by means of heuristic arguments.

For this purpose, it is first useful to recall that the fluctuations of the eigenvalues are characterized by two different scales depending on their location in the spectrum: (i) in the bulk for $\lambda_i/\sqrt{N} = \mathcal{O}(1)$ and $|\lambda_i| < \sqrt{2N}$ and (ii) at the edge where $|\lambda_i \pm \sqrt{2N}| = \mathcal{O}(N^{-1/6})$. The existence of these two scales manifests itself in various observables associated to the eigenvalues of the Gaussian β -ensemble (1), including their mean density defined as:

$$\rho(\lambda, N) = \frac{1}{N} \sum_{i=1}^N \langle \delta(\lambda_i - \lambda) \rangle. \quad (13)$$

One has obviously $\rho(\lambda, N) = \rho(-\lambda, N)$ and one can further show that, for large N , it exhibits two distinct regimes (for $\lambda > 0$) [18, 19, 20, 21]

$$\rho(\lambda, N) \sim \begin{cases} \frac{1}{\sqrt{N}} \rho_{\text{bulk}} \left(\frac{\lambda}{\sqrt{N}} \right), & \lambda < \sqrt{2N} \text{ \& } \lambda = \mathcal{O}(\sqrt{N}), \\ \frac{\sqrt{2}}{N^{5/6}} \rho_{\text{edge}} \left((\lambda - \sqrt{2N}) \sqrt{2} N^{1/6} \right), & |\lambda - \sqrt{2N}| = \mathcal{O}(N^{-1/6}). \end{cases} \quad (14)$$

In Eq. (14), $\rho_{\text{bulk}}(x)$ is the Wigner semi-circle [18, 19]:

$$\rho_{\text{bulk}}(x) = \rho_W(x) = \frac{1}{\pi} \sqrt{2 - x^2}, \quad (15)$$

independently of β , while $\rho_{\text{edge}}(x)$ is given by [20, 21, 22],

$$\rho_{\text{edge}}(x) \sim \begin{cases} [\text{Ai}'(x)]^2 - x \text{Ai}^2(x) + \frac{1}{2} \text{Ai}(x) \left(1 - \int_x^\infty dt \text{Ai}(t) \right), & \beta = 1, \\ [\text{Ai}'(x)]^2 - x \text{Ai}^2(x), & \beta = 2, \\ \kappa^{-1/2} \left([\text{Ai}'(\kappa x)]^2 - \kappa x \text{Ai}^2(\kappa x) - \frac{1}{2} \text{Ai}(\kappa x) \int_{\kappa x}^\infty dt \text{Ai}(t) \right), & \beta = 4, \end{cases} \quad (16)$$

where $\kappa = 2^{2/3}$. Its asymptotic behaviors are given by

$$\begin{cases} \rho_{\text{edge}}(x) \sim \frac{1}{\pi} \sqrt{-x}, & x \rightarrow -\infty, \\ \ln(\rho_{\text{edge}}(x)) \sim -\frac{2\beta}{3} x^{3/2}, & x \rightarrow \infty. \end{cases} \quad (17)$$

Interestingly, one can check that these two regimes for $\rho(\lambda, N)$, the “bulk” one and the “edge” one in Eq. (14), perfectly match when λ approaches the value $\sqrt{2N}$ from below. Indeed, when $\lambda \rightarrow \sqrt{2N}$ from below, $\rho(\lambda, N)$ can be replaced by the Wigner semi-circle (15), which gives:

$$\rho(\lambda, N) \sim \frac{2^{3/4}}{\pi} N^{-3/4} \left(\sqrt{2N} - \lambda \right)^{1/2}, \quad \lambda \rightarrow \sqrt{2N}^-. \quad (18)$$

This behavior (18) coincides with the left tail of the scaling function $\rho_{\text{edge}}(x)$ in Eq. (17). Indeed, when the deviation from $\sqrt{2N}$ is large, $\sqrt{2N} - \lambda \sim \mathcal{O}(\sqrt{N})$, we can substitute in the second line of Eq. (14) the left tail asymptotic behavior of $\rho_{\text{edge}}(x)$ in (17), which gives

$$\rho(\lambda, N) \sim \sqrt{2} N^{-5/6} \frac{1}{\pi} \left(\sqrt{2} N^{1/6} (\sqrt{2N} - \lambda) \right)^{1/2}, \quad \lambda \rightarrow \sqrt{2N}^-, \quad (19)$$

which after a trivial rearrangement coincides with Eq. (18). Note also that the right tail of $\rho_{\text{edge}}(x)$ in Eq. (17) matches, as it should, with the right tail of the TW distribution for β -ensemble [19].

Similarly to the density (14), one expects that, for large N , $\rho_{\text{DOS}}(r, N)$ exhibits two different scaling regimes (see figure 2): (i) a bulk regime, where $r \propto \sqrt{N}$ and (ii) an edge regime where $r = \mathcal{O}(N^{-1/6})$. They can thus be summarized as follows

$$\rho_{\text{DOS}}(r, N) \sim \begin{cases} \frac{1}{\sqrt{N}} \tilde{\rho}_{\text{bulk}}\left(\frac{r}{\sqrt{N}}\right), & c\sqrt{N} < r < 2\sqrt{2N}, \\ \sqrt{2}N^{-5/6} \tilde{\rho}_{\text{edge}}(r\sqrt{2}N^{1/6}), & r = \mathcal{O}(N^{-1/6}), \end{cases} \quad (20)$$

for some real $c < 2\sqrt{2}$ and where $\tilde{\rho}_{\text{bulk}}(x)$ and $\tilde{\rho}_{\text{edge}}(\tilde{r})$ are two different scaling functions.

Let us first investigate the bulk regime. For $r \propto \sqrt{N}$ in (2), one expects that $\rho_{\text{DOS}}(r, N)$ is insensitive to the fluctuations of λ_{max} which are of order $\mathcal{O}(N^{-1/6})$ around the value $\sqrt{2N}$. Therefore, in Eq. (2) the PDF of λ_{max} can be simply replaced by a delta function $\delta(\lambda_{\text{max}} - \sqrt{2N})$. It thus follows that $\rho_{\text{DOS}}(r, N) \approx \rho(\sqrt{2N} - r, N)$ where $\rho(\lambda, N)$ is simply the density of eigenvalues in (13). Therefore, from (14) together with (15) one expects that $\rho_{\text{DOS}}(r, N)$ takes the scaling form given in the first line of Eq. (20) where $\tilde{\rho}_{\text{bulk}}(x)$ is a shifted Wigner semi-circle:

$$\tilde{\rho}_{\text{bulk}}(x) = \rho_W(\sqrt{2} - x) = \frac{1}{\pi} \sqrt{x(2\sqrt{2} - x)}, \quad (21)$$

independently of β .

On the other hand, in the edge regime when $r \sim \mathcal{O}(N^{-1/6})$, the DOS will be sensitive to the fluctuations of λ_{max} and we expect that $\tilde{\rho}_{\text{edge}}(\tilde{r})$ is a non trivial function, as it was shown to be the case for $\beta = 2$ [11] (see also Eq. (24) below). Although we can not compute explicitly this scaling function $\tilde{\rho}_{\text{edge}}(\tilde{r})$ for other values of $\beta \neq 2$ we can extract its asymptotic behaviors for both small and large arguments, which we discuss in the next section.

In Ref. [11] these results were obtained by exact analytical calculations for $\beta = 2$. Here we have performed numerical simulations for different values of $\beta = 1, 2$ and 4 using the tridiagonal random matrices representation [8]. In Fig. 2 we show a plot of $\rho_{\text{DOS}}(r, N)$ as a function of r and different values of N that corroborates the scaling forms in Eq. (20) for three different values of $\beta = 1$ (red), $\beta = 2$ (green) and $\beta = 4$ (blue). The central panel indicates that, in the bulk regime, $\rho_{\text{DOS}}(r, N)$, correctly rescaled, converges to a shifted Wigner law, independently of β . This is in full agreement with

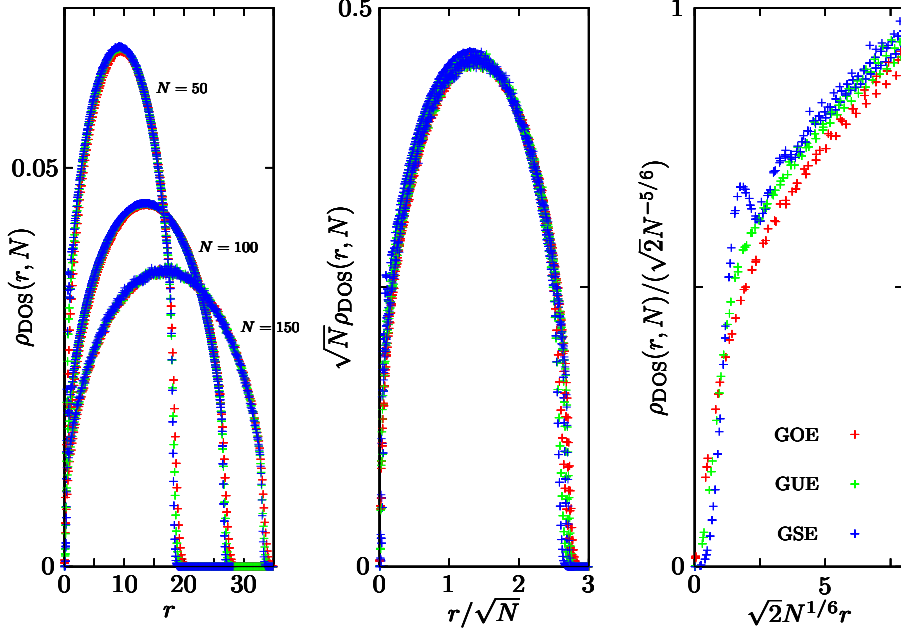


Figure 2. The different scaling regimes for ρ_{DOS} as described in Eq. (20). **Left panel:** Plot of $\rho_{\text{DOS}}(r, N)$ as a function of r for three different values of N and three different values: $\beta = 1$ (red), $\beta = 2$ (green) and $\beta = 4$ (blue). **Central panel:** (bulk regime) Plot of the same quantities (after rescaling) $\sqrt{N} \rho_{\text{DOS}}(r, N)$ as a function of r/\sqrt{N} and for different value of β with the same color code as the one used in the left panel. **Right panel:** (edge regime) Plot of the same quantities (after yet another rescaling) $\rho_{\text{DOS}}(r, N)/(\sqrt{2} N^{-5/6})$ as a function of $\sqrt{2} N^{1/6} r$.

the scaling form in the first line of Eq. (20). Finally, the right panel shows a plot of $\rho_{\text{DOS}}(r, N)$ for small r , which is in a good agreement with the scaling form given in the second line of Eq. (20). It also indicates that the limiting function $\tilde{\rho}_{\text{edge}}(\tilde{r})$ depends explicitly on β .

We conclude this section by mentioning that the typical fluctuations of the first gap are naturally expected to scale as $d_{1,N} = \Lambda_{1,N} - \Lambda_{2,N} \sim \mathcal{O}(N^{-1/6})$, as the fluctuations of $\Lambda_{1,N} = \lambda_{\text{max}}$ and $\Lambda_{2,N}$ around $\sqrt{2N}$ are also of order $\mathcal{O}(N^{-1/6})$ [1, 2]. Hence, for large N we expect that $p_{\text{GAP}}(r, N)$ takes the scaling form:

$$p_{\text{GAP}}(r, N) = \sqrt{2} N^{1/6} \tilde{p}_{\text{typ}} \left(r \sqrt{2} N^{1/6} \right), \quad (22)$$

where the factor $\sqrt{2}$ in the argument has been chosen here for convenience,

according to the choice made for $\rho_{\text{DOS}}(r, N)$ in the second line of Eq. (20). This scaling form (22) together with the scaling function $\tilde{p}_{\text{typ}}(\tilde{r})$ were obtained exactly for $\beta = 2$. It is natural to expect that this scaling form (22) holds for any value of $\beta > 0$, with a β -dependent scaling function $\tilde{p}_{\text{typ}}(\tilde{r})$, and below we give some heuristic arguments to obtain the asymptotic behaviors of the scaling function $\tilde{p}_{\text{typ}}(\tilde{r})$.

4. Asymptotic behaviors of the scaling functions $\tilde{\rho}_{\text{edge}}(\tilde{r})$ and $\tilde{p}_{\text{typ}}(\tilde{r})$

We first begin to analyze the small \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$. From the definition of the DOS in Eq. (2) it is clear that its small \tilde{r} behavior is directly related to the probability that two eigenvalues – namely the first one and the second one – become extremely close to each other. From the joint PDF of the eigenvalues (1), this probability vanishes as \tilde{r}^β as a consequence of the short distance repulsion between eigenvalues, which comes from the Vandermonde term $\prod_{i,j} |\lambda_i - \lambda_j|^\beta$ in the joint PDF (1). One thus expects that $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta$, with an a priori unknown constant a_β .

On the other hand, it is reasonable to assume that there is a smooth matching between the edge region and the bulk region described by the shifted Wigner semi-circle law (21), as it is the case for the density of eigenvalues [see the discussion between Eqs. (18) and (19)]. This means that the large \tilde{r} behavior of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ has to coincide with the small argument of the shifted Wigner semi circle (21), which does not depend on β . Hence one deduces that $\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \sqrt{\tilde{r}}/\pi$, for all β . The asymptotic behaviors of the DOS in the edge regime can thus be summarized as follows:

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) \sim \begin{cases} a_\beta \tilde{r}^\beta + o(\tilde{r}^\beta), & \tilde{r} \rightarrow 0, \\ \frac{\sqrt{\tilde{r}}}{\pi} + o(\tilde{r}^{1/2}), & \tilde{r} \rightarrow \infty. \end{cases} \quad (23)$$

These behaviors (23) have been confirmed, for $\beta = 2$, by an exact calculation of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ [11]:

$$\tilde{\rho}_{\text{edge}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}(\tilde{r}, u)^2 - \left(\int_x^{\infty} q(u) \tilde{f}(\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx, \quad (24)$$

where $\mathcal{F}_2(x)$ is the Tracy-Widom distribution associated to GUE [1] and $q(x)$ is the Hastings-McLeod solution of the Painlevé II equation

$$\mathcal{F}_2(x) = \exp \left[- \int_x^{\infty} (u - x) q^2(u) du \right], \quad \begin{cases} q''(x) = 2q^3(x) + x q(x), \\ q(x) \sim \text{Ai}(x) \text{ for } x \rightarrow \infty, \end{cases} \quad (25)$$

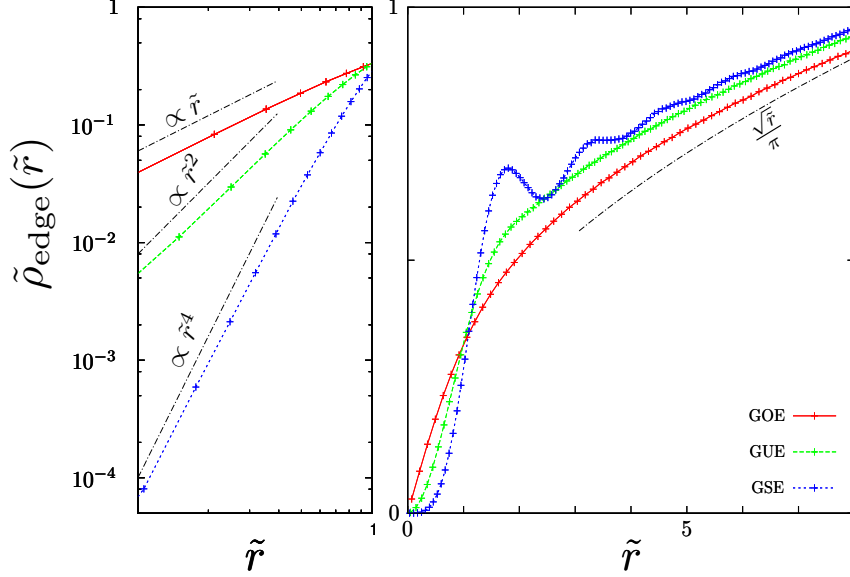


Figure 3. Plot of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ for three different values of β ($\beta = 1$ in red, $\beta = 2$ in green and $\beta = 4$ in blue). **Left panel:** the data are plotted on a log-log scale and they exhibit a small \tilde{r} behavior compatible with our heuristic arguments in (23). **Right panel:** plot of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ in terms of \tilde{r} with the asymptotic behavior for large \tilde{r} compatible with our heuristic arguments (23).

while $\tilde{f}(\tilde{r}, x)$ satisfies

$$\begin{cases} \partial_x^2 \tilde{f}(\tilde{r}, x) - [x + 2q^2(x)] \tilde{f}(\tilde{r}, x) = -\tilde{r} \tilde{f}(\tilde{r}, x) , \\ \tilde{f}(\tilde{r}, x) \underset{x \rightarrow \infty}{\sim} 2^{-1/6} \sqrt{\pi} \text{Ai}(x - \tilde{r}) . \end{cases} \quad (26)$$

It was further shown in [11] that $\tilde{f}(\tilde{r}, x)$ can be expressed in terms of the Lax pair associated to the Painlevé XXXIV equation. From this exact expression (24) one can derive the asymptotic behaviors announced in (23) and compute explicitly the amplitude $a_2 = 1/2$ [11].

In Fig. 3 we show a plot of $\tilde{\rho}_{\text{edge}}(\tilde{r})$ as a function of \tilde{r} for three different values of $\beta = 1, 2$ and 4 exhibiting the small \tilde{r} behavior (in the left panel) and the large \tilde{r} behavior (in the right panel). In each case, $\tilde{\rho}_{\text{edge}}(\tilde{r})$ was computed by sampling 2.10^7 independent random matrices of size 100×100 . Both panels show a good agreement with our predictions in (23). We also notice on the right panel of Fig. 3 that $\tilde{\rho}_{\text{edge}}(\tilde{r})$ exhibits clear oscillations for $\beta = 4$, while the curves for $\beta = 1$ and 2 are much smoother.

We now turn to the small \tilde{r} asymptotic behavior of the gap distribution $\tilde{p}_{\text{typ}}(\tilde{r})$, which can be argued to coincide with the small \tilde{r} behavior of the DOS, $\tilde{\rho}_{\text{edge}}(\tilde{r})$. Indeed, for small \tilde{r} , the contribution to $\tilde{\rho}_{\text{edge}}(\tilde{r})$ comes only from the gap between the two largest eigenvalues because the others (the third, the fourth etc. eigenvalues) are too far, as a consequence of the short range repulsion coming from the Vandermonde term in Eq. (1). Hence one expects that $\tilde{p}_{\text{typ}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta$ with the same amplitude a_β as above in (23).

On the other hand, the large \tilde{r} behavior of $\tilde{p}_{\text{typ}}(\tilde{r})$ can be obtained, to leading order, by assuming that for large separation \tilde{r} the two first eigenvalues $\Lambda_{2,N}$ and $\Lambda_{1,N} = \lambda_{\text{max}}$ become statistically independent. Hence one expects:

$$\begin{aligned} p_{\text{GAP}}(r, N) &= \int_{-\infty}^{\infty} \text{Pr}[\Lambda_{2,N} = \lambda, \lambda_{\text{max}} = \lambda + r] d\lambda \\ &\underset{r \rightarrow \infty}{\sim} \int_{-\infty}^{\infty} \text{Pr}[\Lambda_{2,N} = \lambda] \text{Pr}[\lambda_{\text{max}} = \lambda + r] d\lambda \\ &\underset{r \rightarrow \infty}{\sim} \text{Pr}[\lambda_{\text{max}} = r] . \end{aligned} \quad (27)$$

Therefore, we expect that right tail of $\tilde{p}_{\text{typ}}(\tilde{r})$ coincides with the right tail of the TW distribution which goes for large argument as $\ln \mathcal{F}'_\beta(x) \sim -\frac{2\beta}{3}x^{3/2}$. To summarize, we obtain the asymptotic behaviors of $\tilde{p}_{\text{typ}}(\tilde{r})$ as:

$$\begin{cases} \tilde{p}_{\text{typ}}(\tilde{r}) \sim a_\beta \tilde{r}^\beta + o(\tilde{r}^\beta) & \tilde{r} \rightarrow 0 \\ \ln \tilde{p}_{\text{typ}}(\tilde{r}) \sim -\frac{2\beta}{3}\tilde{r}^{3/2} + o(\tilde{r}^{3/2}) , & \tilde{r} \rightarrow +\infty , \end{cases} \quad (28)$$

while the computation of a_β as well as the subleading corrections, both for small and large arguments, for any β remains challenging.

In Ref. [11], we obtained an exact expression of $\tilde{p}_{\text{typ}}(\tilde{r})$ for the special case $\beta = 2$ as

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \frac{2^{1/3}}{\pi} \int_{-\infty}^{\infty} \left[\tilde{f}^2(-\tilde{r}, x) - \left(\int_x^{\infty} q(u) \tilde{f}(-\tilde{r}, u) du \right)^2 \right] \mathcal{F}_2(x) dx . \quad (29)$$

From this exact formula (29) we could not only check the above asymptotic behaviors (28) but also obtain the sub-leading terms as:

$$\tilde{p}_{\text{typ}}(\tilde{r}) = \begin{cases} \frac{1}{2}\tilde{r}^2 + a_4\tilde{r}^4 + \mathcal{O}(\tilde{r}^6) \\ A \exp\left(-\frac{4}{3}\tilde{r}^{3/2} + \frac{8}{3}\sqrt{2}\tilde{r}^{3/4}\right) \tilde{r}^{-\frac{21}{32}} \left(1 - \frac{1405\sqrt{2}}{1536}\tilde{r}^{-3/4} + \mathcal{O}(\tilde{r}^{-3/2})\right) , \end{cases} \quad (30)$$

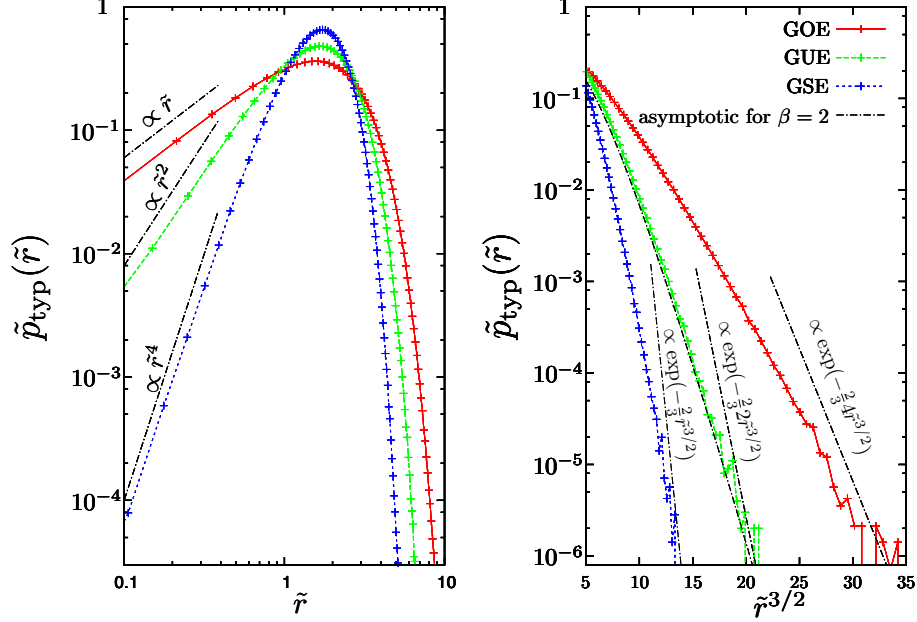


Figure 4. Comparison of $\tilde{p}_{\text{typ}}(\tilde{r})$ for the different β -ensemble with $2 \cdot 10^7$ realizations for a 100×100 matrix for each case. ($\beta = 1$ in red, $\beta = 2$ in green and $\beta = 4$ in blue) **Left panel:** log-log plot of $\tilde{p}_{\text{typ}}(\tilde{r})$ with the asymptotic behavior for small \tilde{r} . **Right panel:** log plot of $\tilde{p}_{\text{typ}}(\tilde{r})$ in terms of $\tilde{r}^{3/2}$ with the asymptotic behavior for large \tilde{r} . We also plot the asymptotic behavior for large \tilde{r} obtained in the $\beta = 2$ case Eq. (30).

where the first and second lines correspond respectively to the small and large \tilde{r} behaviors. In (30), the amplitude A is given by $A = 2^{-91/48} e^{\zeta'(-1)} / \sqrt{\pi}$, where $\zeta'(x)$ is the derivative of the Riemann zeta function, while a_4 can be expressed in terms of integrals involving $q(x)$ with the result $a_4 \sim -0.393575\dots$. It should be noticed that a different expression, somehow more complicated, of the PDF of the first gap $\tilde{p}_{\text{typ}}(\tilde{r})$ had been obtained previously in Ref. [23], involving also Painlevé transcendents. It remains an open question to show that these two expressions do coincide.

5. The density of states and the gap for a fixed value of λ_{max}

Up to now, we have studied the DOS $\rho_{\text{DOS}}(r, N)$ and the PDF of the gap $p_{\text{GAP}}(r, N)$ averaged over the value of λ_{max} . It is also interesting to look at these quantities for fixed value of $\lambda_{\text{max}} = y$, the corresponding quantities being denoted as $\rho_{\text{DOS}}(r|y, N)$ and $p_{\text{GAP}}(r|y, N)$. We naturally expect the

scaling forms, valid for all $\beta > 0$:

$$\rho_{\text{DOS}}(r|y, N) = N^{-5/6} \sqrt{2} \tilde{\rho}_{\text{edge}}(\sqrt{2}N^{1/6}r|\sqrt{2}N^{1/6}(y - \sqrt{2N})), \quad (31)$$

$$p_{\text{GAP}}(r|y, N) = N^{1/6} \sqrt{2} \tilde{p}_{\text{typ}}(\sqrt{2}N^{1/6}r|\sqrt{2}N^{1/6}(y - \sqrt{2N})). \quad (32)$$

From the results of Ref. [11] one can compute explicitly these scaling functions as:

$$\begin{aligned} \tilde{\rho}_{\text{edge}}(\tilde{r}|x) = \frac{2^{1/3}}{\pi} \mathcal{F}_2(x) & \left[R(x) \left(\left(\tilde{r} + \frac{R(x)}{q^2(x)} \right) \tilde{f}^2(\tilde{r}, x) - 2 \frac{q'(x)}{q(x)} \tilde{f}(\tilde{r}, x) \tilde{g}(\tilde{r}, x) \right. \right. \\ & \left. \left. + \left(1 + \frac{q^2(x)}{\tilde{r}} \right) \tilde{g}(\tilde{r}, x)^2 \right) - \frac{q^2(x)}{\tilde{r}^2} \tilde{g}^2(\tilde{r}, x) \right], \quad (33) \end{aligned}$$

with $R(x) = \int_x^\infty q^2(u) du$ and $q(x) \tilde{g}(\tilde{r}, x) = -\tilde{r} \int_x^\infty q(u) \tilde{f}(\tilde{r}, u) du$. Similarly, for the gap one finds:

$$\begin{aligned} \tilde{p}_{\text{typ}}(\tilde{r}|x) = \frac{2^{1/3}}{\pi} \mathcal{F}_{2,s} & \left[R_s \left(\left(-\tilde{r} + \frac{R_s}{q_s^2} \right) \tilde{f}_s^2 - 2 \frac{q'_s}{q_s} \tilde{f}_s \tilde{g}_s + \left(1 - \frac{q_s^2}{\tilde{r}} \right) \tilde{g}_s^2 \right) \right. \\ & \left. - \frac{q_s^2}{\tilde{r}^2} \tilde{g}_s^2 \right], \quad (34) \end{aligned}$$

where we have now used the shorthand notations for the shifted quantities (hence the subscript s) $q_s = q(x - \tilde{r})$, $R_s = R(x - \tilde{r})$, $\mathcal{F}_{2,s} = \mathcal{F}_2(x - \tilde{r})$, $\tilde{f}_s = \tilde{f}(-\tilde{r}, x - \tilde{r})$, $\tilde{g}_s = \tilde{g}(-\tilde{r}, x - \tilde{r})$. In Fig. (5) we show a plot of the constrained quantities: $\tilde{\rho}_{\text{edge}}(\tilde{r}|0)$ in the left panel and $\tilde{p}_{\text{typ}}(\tilde{r}|0)$ in the right panel when $\lambda_{\text{max}} = \sqrt{2N}$ ie $x = 0$ for three different values of β ($\beta = 1$ in red, $\beta = 2$ in green and $\beta = 4$ in blue) computed by sampling 2.10^7 independent tridiagonal random matrices of size 100×100 [8] for each β and only kept the events when $|\lambda_{\text{max}} - \sqrt{2N}| < 0.1$. Both panels show a good agreement with our predictions in (33, 34) for $\beta = 2$ (green solid line) [11].

6. Conclusion

To conclude, we have studied the phenomenon of “crowding” of the eigenvalues near the largest eigenvalue λ_{max} for random matrices belonging to the Gaussian β -ensembles (1). In particular we focused on the DOS (2), which is the density of eigenvalues seen from the largest eigenvalue, and the PDF of the first gap, these two quantity being related for any finite N through the relation (4). Based on exact results obtained for Hermitian matrices, corresponding to $\beta = 2$, obtained in [11], and using heuristic arguments

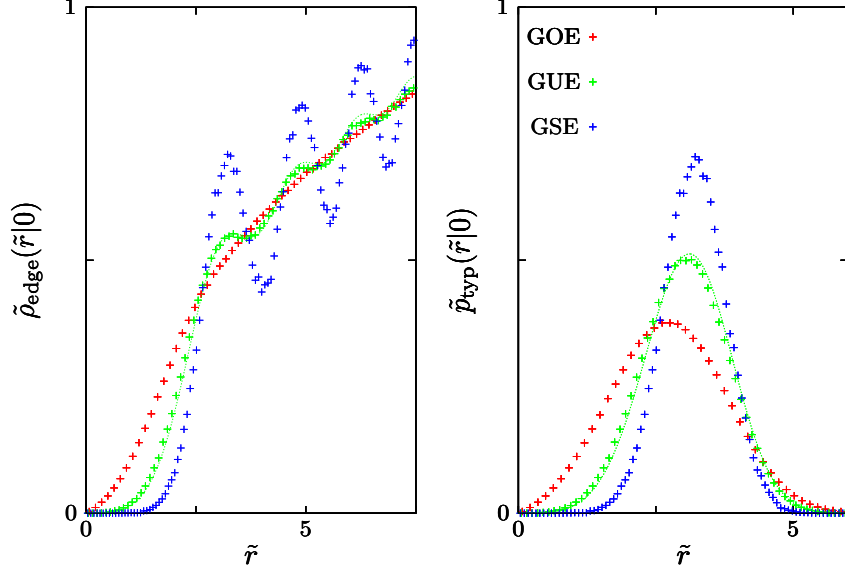


Figure 5. Comparison of $\tilde{\rho}_{\text{edge}}(\tilde{r}|0)$ (**left panel**) and $\tilde{p}_{\text{typ}}(\tilde{r}|0)$ (**right panel**) for three different values of β ($\beta = 1$ in red, $\beta = 2$ in green and $\beta = 4$ in blue). We have performed 2.10^7 realizations of a 200×200 matrix. For each case, we keep only events when $|\lambda_{\text{max}} - \sqrt{2N}| < 0.1$ (dots). We also plot in solid line the exact result for $\beta = 2$ (33) and (34) obtained in [11] (using the numerical values of $q(x)$ and $\mathcal{F}_2(x)$ from Ref. [24]).

we obtained a general description of these two quantities in the large N limit. We also presented results of numerical simulations supporting our arguments. As we have seen in section 2, these quantities $\rho_{\text{DOS}}(r, N)$ and $p_{\text{GAP}}(r, N)$ enter naturally into the computation of physical observables in the fully connected spherical spin-glass model and it will be particularly interesting to explore further the implications of our results to the relaxational dynamics of this model. Another application of the techniques developed in [11], which provided a detailed analysis of an orthogonal polynomial system initially introduced in [12], concerns the level curvature distribution at the soft edge of random Hermitian matrices which also involves the same orthogonal polynomials [25]. Finally, it will be interesting to extend the present study to other ensembles of random matrices, like the Laguerre-Wishart ensemble. Hence we hope that these results will motivate further studies of

near extreme eigenvalues.

Acknowledgments. We acknowledge support by ANR grant 2011-BS04-013-01 WALKMAT and in part by the Indo-French Centre for the Promotion of Advanced Research under Project 4604 – 3. GS acknowledges support from Labex-PALM (Project Randmat). We also thank Y. V. Fyodorov for useful discussions.

References

- [1] C. A. Tracy, H. Widom, *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys. **159**, 151 (1994).
- [2] C. A. Tracy, H. Widom, *On orthogonal and symplectic matrix ensembles*, Commun. Math. Phys. **177**, 727 (1996).
- [3] G. Ben Arous, A. Dembo, A. Guionnet, *Aging of spherical spin glasses*, Probab. Theory Relat. Fields **120**, 1 (2001).
- [4] D. S. Dean, S. N. Majumdar, *Large deviations of extreme eigenvalues of random matrices*, Phys. Rev. Lett. **97**, 160201 (2006).
- [5] D. S. Dean, S. N. Majumdar, *Extreme value statistics of eigenvalues of Gaussian random matrices*, Phys. Rev. E **77**, 041108 (2008).
- [6] S. N. Majumdar, M. Vergassola, *Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices*, Phys. Rev. Lett. **102**, 060601 (2009).
- [7] S. N. Majumdar, G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, J. Stat. Mech. P01012 (2014).
- [8] I. Dumitriu, A. Edelman, *Matrix models for beta ensembles*, J. Math. Phys. **43**, 5830 (2002).
- [9] S. Sabhapandit, S. N. Majumdar, *Density of near-extreme events*, Phys. Rev. Lett. **98**(14), 140201 (2007).
- [10] A. Perret, A. Comtet, S. N. Majumdar, G. Schehr, *Near-Extreme Statistics of Brownian Motion*, Phys. Rev. Lett. **111**, 240601 (2013).
- [11] A. Perret, G. Schehr, *Near-Extreme Eigenvalues and the First Gap of Hermitian Random Matrices*, J. Stat. Phys. 156:843-876 (2014).
- [12] C. Nadal, S. N. Majumdar, *A simple derivation of the Tracy-Widom distribution of the maximal eigenvalue of a Gaussian unitary random matrix*, J. Stat. Mech., P04001 (2011).
- [13] Y. V. Fyodorov, private communication after a talk given by one of us at the conference *Random Matrix Theory : Foundations and Applications*, Kraków, (2014).
- [14] C. Monthus, T. Garel, *Typical versus averaged overlap distribution in Spin-Glasses: Evidence for the droplet scaling theory*, Phys. Rev. B **88**, 134204 (2013).

- [15] L. F. Cugliandolo, D. S. Dean, *Full dynamical solution for a spherical spin-glass model*, J. Phys. A: Math. Gen. **28**, 4213 (1995).
- [16] J. Kurchan, L. Laloux, *Phase space geometry and slow dynamics*, J. Phys. A: Math. Gen. **29**, 1929 (1996).
- [17] Y. V. Fyodorov, A. Perret, G. Schehr, in preparation.
- [18] M. L. Mehta, *Random Matrices*, 2nd Edition, Academic Press (1991).
- [19] P. J. Forrester, *Log-gases and random matrices*, Princeton University Press, Princeton, NJ, (2010).
- [20] M. Bowick, E. Brézin, *Universal scaling of the tail of the density of eigenvalues in random matrix models*, Phys. Lett. B **268**, 21 (1991).
- [21] P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nucl. Phys. B **402**(3), 709 (1993).
- [22] P. J. Forrester, N. E. Frankel, T. M. Garoni, *Asymptotic form of the density profile for Gaussian and Laguerre random matrix ensembles with orthogonal and symplectic symmetry*, J. Math. Phys. **47**(2), 023301 (2006).
- [23] N. S. Witte, F. Bornemann, P. J. Forrester, *Joint distribution of the first and second eigenvalues at the soft edge of unitary ensembles*, Nonlinearity **26**, 1799 (2013).
- [24] This was done using the tabulation of $q(x)$, $R(x)$ and $\mathcal{F}_2(x)$ which can be found on the webpage of M. Prähofer: <http://www-m5.ma.tum.de/KPZ>.
- [25] Y. V. Fyodorov, *Level curvature distribution: from bulk to the soft edge of random Hermitian matrices*, Acta Phys. Pol. A **120**(6), 100 (2012).

On certain functionals of the maximum of Brownian motion and their applications.

[123] A. Perret, A. Comtet, S.N. Majumdar, G. Schehr (2015).

Prépublication arXiv :1502.01218

Soumis à Journal of Statistical Physics

Anthony Perret · Alain Comtet ·
Satya N. Majumdar · Grégory Schehr

On certain functionals of the maximum of Brownian motion and their applications

January 28, 2015

Abstract We consider a Brownian motion (BM) $x(\tau)$ and its maximal value $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$ on a fixed time interval $[0, t]$. We study functionals of the maximum of the BM, of the form $\mathcal{O}_{\max}(t) = \int_0^t V(x_{\max} - x(\tau)) d\tau$ where $V(x)$ can be any arbitrary function and develop various analytical tools to compute their statistical properties. These tools rely in particular on (i) a “counting paths” method and (ii) a path-integral approach. In particular, we focus on the case where $V(x) = \delta(x - r)$, with r a real parameter, which is relevant to study the density of near-extreme values of the BM (the so called density of states), $\rho(r, t)$, which is the local time of the BM spent at given distance r from the maximum. We also provide a thorough analysis of the family of functionals $T_\alpha(t) = \int_0^t (x_{\max} - x(\tau))^\alpha d\tau$, corresponding to $V(x) = x^\alpha$, with α real. As α is varied, $T_\alpha(t)$ interpolates between different interesting observables. For instance, for $\alpha = 1$, $T_{\alpha=1}(t)$ is a random variable of the “area”, or “Airy”, type while for $\alpha = -1/2$ it corresponds to the maximum time spent by a ballistic particle through a Brownian random potential. On the other hand, for $\alpha = -1$, it corresponds to the cost of the optimal algorithm to find the maximum of a discrete random walk, proposed by Odlyzko. We revisit here, using tools of theoretical physics, the statistical properties of this algorithm which had been studied before using probabilistic methods. Finally, we extend our methods to constrained BM, including in particular the Brownian bridge, i.e., the Brownian motion starting and ending at the origin.

Keywords Brownian motion · Extreme statistics · Path integral

A. Perret
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France

A. Comtet
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France, Université Pierre et Marie Curie, Paris 6, 75252 Paris Cedex 05, France

S. N. Majumdar
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France

G. Schehr
Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100, 91405 Orsay Cedex, France

1 Introduction

Stochastic processes are at the heart of many fundamental problems in statistical physics. In particular, it was realized a long time ago that Brownian motion (BM) is the process underlying many physical systems and corresponding random models. Since then, BM has not only become a cornerstone of statistical physics [1, 2, 3] but has also found numerous applications in various areas of science, including biology [4], computer science [5, 6, 7, 8, 9] or financial mathematics [10, 11]. These various applications have motivated the study of functionals of Brownian motion [8, 10, 12, 13], which are observables of the form $\mathcal{O}(t) = \int_0^t V(x(\tau)) d\tau$ where $V(x)$ can be any function and $x(\tau)$ is a BM (see Fig. 1 (a)). It might also be relevant to consider functionals of variants of BM, like the Brownian bridge (BB), $x_B(\tau)$, which is a BM conditioned to start and end at the origin (see Fig. 1 (b)), the Brownian excursion (BE), $x_E(\tau)$, which is a BB conditioned to stay positive on the whole time interval $[0, t]$ (see Fig. 1 (c)) as well as the Brownian meander which is constrained to stay positive on $[0, t]$ but can end up at any point at time t (see Fig. 1 (d)). For instance, if $V(x) = \delta(x - a)$, the Brownian functional $\mathcal{O}(t)$ corresponds to the local time at the fixed level a , which is an important quantity in probability theory [14]. Another interesting example concerns the case where $V(x) = x$ for a BE, which corresponds to the area under a BE. In this case, the distribution of $\mathcal{O}_E(t) = \int_0^t x_E(\tau) d\tau$ is given by the so called Airy-distribution which appears in computer science [15, 16, 17, 18] as well as in the extreme statistics of elastic interfaces [8, 19, 20]. Extensions of the Airy-distribution to the area under Bessel processes (*i.e.*, radius of the d -dimensional process) have been recently discussed in Ref. [21]. Yet another example which is relevant in finance is the case where $V(x) = \exp(x)$, which describes the price of an Asian stock option in the Black-Scholes framework [22]. Note that in this exponential case, $\mathcal{O}(t) = \int_0^t \exp(x(\tau)) d\tau$ also represents the stationary current of a disordered Sinai chain connected to two reservoirs of particles [23, 24, 25, 26]. Quite interestingly, these functionals of BM and its variants can be studied using powerful tools of theoretical physics, namely path integrals methods (leading to the so called Feynman-Kac formula). This allows to recast the study of Brownian functionals in a quantum mechanical framework [8, 27].

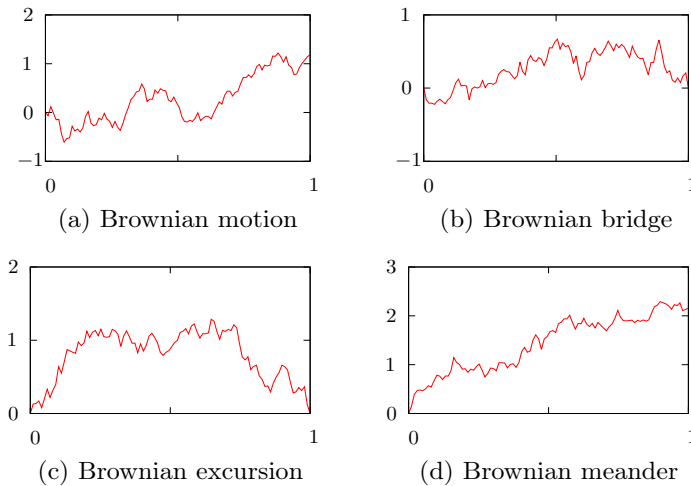


Fig. 1 Brownian motion $x(\tau)$ (a) and its variants: the Brownian bridge $x_B(\tau)$ in (b), the Brownian excursion $x_E(\tau)$ in (c) and the Brownian meander $x_{Me}(\tau)$ in (d).

While such Brownian functionals are thus well understood, several recent works have pointed out the physical relevance of *functionals of the maximum of Brownian motion*, which are much less studied. In this case, one is interested in observables of the form $\mathcal{O}_{\max}(t) = \int_0^t V(x_{\max} - x(\tau)) d\tau$ with $x_{\max} = \max_{0 \leq \tau \leq t} x(\tau)$ where $x(\tau)$ is BM or one of its variants (see Fig. 1). An important case

corresponds to the case where $V(x) = \delta(x - r)$ where

$$\rho(r, t) = \int_0^t \delta(x_{\max} - x(\tau) - r) d\tau, \quad (1)$$

which is the so called density of states (DOS) near the maximum. This is a natural and useful quantity to characterize the crowding of near-extremes [28]. Indeed, $\rho(r, t)dr$ denotes the amount of time spent by $x(\tau)$ at a distance within the interval $[r, r + dr]$ from x_{\max} (see Fig. 2). Hence $\rho(r, t)$ is similar to the local time with the major difference that here the distances are measured from x_{\max} , which is itself a random variable. The statistics of the DOS was recently studied by us in the context of near-extreme statistics [29]. Note that, by definition, $\int_0^\infty \rho(r, t) dr = t$. Therefore its average value, $\langle \rho(r, t) \rangle / t$, where $\langle \dots \rangle$ means an average over the trajectories of BM has a natural probabilistic interpretation as it is the probability density function to find the BM at a given distance r from the maximum in the time interval $[0, t]$. In particular, the average value of any functional of the maximum can be expressed as

$$\langle \mathcal{O}_{\max}(t) \rangle = \langle \int_0^t V(x_{\max} - x(\tau)) d\tau \rangle = \int_0^\infty \langle \rho(r, t) \rangle V(r) dr, \quad (2)$$

which naturally holds not only for BM but also for its variants, like Brownian bridge or Brownian excursion.

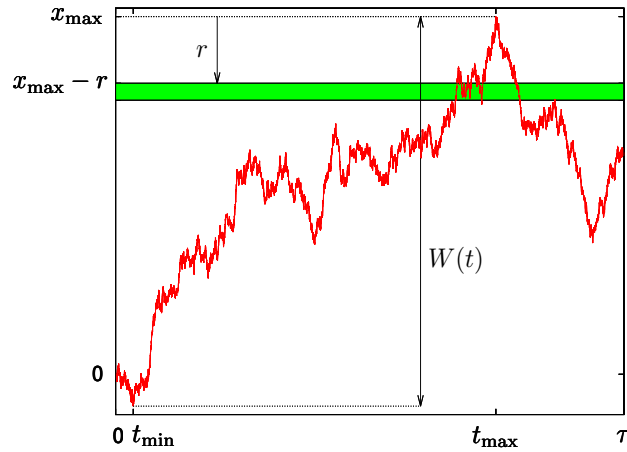


Fig. 2 (Color online) One realization of the process $x(\tau)$ on the time interval $[0, t]$, with a width $W(t) = \max_{\tau \in [0, t]} x(\tau) - \min_{\tau \in [0, t]} x(\tau)$. $x(\tau)$ spends a time $\rho(r, t)dr$ at a distance within $[r, r + dr]$ (the green stripe) from the maximum x_{\max} , with $\rho(r, t)$ being the DOS (1).

Another very interesting application of functionals of the maximum of Brownian motion concerns the case where $V(x) = 1/(2x)$, which enters into the analysis of the optimal algorithm to find the maximum of a discrete random walk of n steps with $n \gg 1$ [30]. Indeed, let us consider a discrete random walk (RW), starting from $X_0 = 0$ and evolving via the Markov rule: $X_k = X_{k-1} + \eta_k$ where $\eta_k = \pm 1$ with equal probability $1/2$. We study the search problem of finding the maximum of the RW, $M_n = \max_{0 \leq i \leq n} X_i$, while minimizing the number of values of X_k 's that are probed. The cost of the algorithm is identified by the number of probes used to find M_n . The simplest algorithm consists of probing all positions X_k for $1 \leq k \leq n$: its cost is n . In Fig. 12, we show an example for $n = 14$ where we find $M_{14} = 7$ in 4 probes: this shows that, because of strong correlations between the positions of the RW, n is actually a rough upper bound of the cost of the optimal algorithm. Of course, some RWs need more probes than others in order to find M_n . For example, the RW with n jumps $+1$ needs only one probe in $X_n = n$ but the RW with alternating jumps ± 1 ($X_{2k} = 0$ and $X_{2k+1} = 1$) needs $\lfloor \frac{n}{2} \rfloor + 1$ probes to be sure that none $X_{2k} = 2$. But of course these cases are rare. Let A_n be the ensemble of algorithms that find M_n , according to the above rules, and let us denote by $C(a)$ the cost, as defined above, of the algorithm a that belongs to A_n . Of course $C(a)$ is a random variable, which

varies from one realization of the RW to another, its average value being denoted by $\langle C(a) \rangle$. In Ref. [30], Odlyzko studied the minimal average cost of such algorithms and he showed that, for large n , the minimal average cost is proportional to \sqrt{n} , much smaller than the cost of the aforementioned naive algorithm, which necessitates n probes. Indeed, one has [30]

$$\min_{a \in A_n} \langle C(a) \rangle = c_0 \sqrt{n} + o(\sqrt{n}) , \quad (3)$$

where c_0 is a constant given by

$$c_0 = \langle I \rangle , \quad I = \frac{1}{2} \int_0^1 \frac{d\tau}{[x_{\max} - x(\tau)]} , \quad (4)$$

hence the relevance of the functional of the maximum with $V(x) = 1/(2x)$ for this search problem. Note that Odlyzko found an expression of c_0 in terms of a complicated double integral which was then evaluated independently by Hwang [31] and Chassaing [32] to be

$$c_0 = \sqrt{\frac{8}{\pi}} \log 2 = 1.1061 \dots . \quad (5)$$

Furthermore, Odlyzko found an algorithm, denoted as $\text{Od}(n)$, belonging to A_n , such that

$$\langle C(\text{Od}(n)) \rangle = c_0 \sqrt{n} + o(\sqrt{n}) . \quad (6)$$

Hence Eqs. (3) and (6) show that Odlyzko's algorithm $\text{Od}(n)$ is on average asymptotically optimal. In a subsequent work, Chassaing, Marckert and Yor showed that $\text{Od}(n)$ is not only asymptotically optimal on average (6) but also in distribution [33], which means the following. If one defines $\Phi_n(x)$ as

$$\Phi_n(x) = \min_{a \in A_n} \Pr \left(\frac{C(a)}{\sqrt{n}} \geq x \right) , \quad (7)$$

then one has for any x [33]:

$$\lim_{n \rightarrow \infty} \Phi_n(x) = \Pr(I \geq x) , \quad (8)$$

where I is the random variable defined above in (4). In Ref. [33], the authors studied the distribution of I as well as its moments, using rather involved probabilistic methods. We will show here how these results can be derived simply using path integrals techniques.

Yet another case of a functional of the maximum, $\mathcal{O}_{\max}(t)$, corresponds to the case where $V(x) \propto 1/\sqrt{x}$. In this case $\mathcal{O}_{\max}(t)$ describes the largest exit time of a particle, of unit mass, moving ballistically through a random potential on the segment $[0, 1]$ (at zero temperature). Consider a random Brownian potential $x(y)$ over a line segment $y \in [0, 1]$. Imagine shooting a classical particle of unit mass with fixed energy E from the left of the segment at $y = 0$. The energy conservation leads to

$$\frac{1}{2} \left(\frac{dy}{dt} \right)^2 + x(y) = E . \quad (9)$$

Clearly, this classical particle can penetrate the region $y \in [0, 1]$ if and only if its energy E is bigger than the maximum value of the potential $x(y)$ over $y \in [0, 1]$, i.e., if $E > x_{\max}$ where $x_{\max} = \max_{0 \leq y \leq 1} [x(y)]$. Now, imagine sending a beam of classical particles with varying energy through this potential barrier. Only those particles with energy larger than x_{\max} will go through the barrier. The time taken for such a penetrating particle to exit the region $y \in [0, 1]$ through its right can be computed from Eq. (9) as

$$\mathcal{T}(E) = \frac{1}{\sqrt{2}} \int_0^1 \frac{dy}{\sqrt{E - x(y)}} \quad (10)$$

where $E \geq x_{\max}$. Clearly, the time needed by a penetrating particle to cross the region $y \in [0, 1]$ depends on the energy E of the particle. The slowest particle, i.e., the one that takes the longest time

to cross, is the one that has the lowest allowed energy to penetrate, i.e., the one with energy $E = x_{\max}$. Hence, the maximum time needed by a particle to cross the barrier is given by

$$\mathcal{T}_{\max} = \max_{E \geq x_{\max}} [\mathcal{T}(E)] = [\mathcal{T}(x_{\max})] = \frac{1}{\sqrt{2}} \int_0^1 \frac{dy}{\sqrt{x_{\max} - x(y)}} \quad (11)$$

which thus corresponds to a functional of the maximum of BM, $\mathcal{O}_{\max}(t=1)$, with $V(x) = 1/\sqrt{2x}$. In view of these two interesting physical examples in Eqs. (4) and (11), it is rather natural to consider the family of functionals of the maximum of BM $\mathcal{O}_{\max}(t)$ with $V(x) = x^\alpha$ such that

$$T_\alpha(t) = \int_0^t d\tau (x_{\max} - x(\tau))^\alpha, \quad (12)$$

indexed by a real $\alpha \in]-2, \infty[$. Note that, by using the self-affinity of BM, $T_\alpha(t) \stackrel{\text{law}}{=} t^{1+\alpha/2} T_\alpha(t=1)$. In particular for $\alpha = -1$ this corresponds to I in (4) while for $\alpha = -1/2$ this corresponds to \mathcal{T}_{\max} in (11). On the other hand, for $\alpha = 1$, $T_{\alpha=1}(t)$ is called the area under a Brownian double meander [20, 18]. For other values of α , $T_\alpha(t)$ generalizes these three cases.

Of course, one can consider similar observables as in (12) for the Brownian bridge, i.e. when $x(\tau) \rightarrow x_B(\tau)$ (see Fig. 1 (b)) and $x_{\max,B}$ denotes its maximum on the time interval $[0, t]$:

$$T_{\alpha,B}(t) = \int_0^t (x_{\max,B} - x_B(\tau))^\alpha d\tau, \quad (13)$$

with, as above, $T_{\alpha,B}(t) \stackrel{\text{law}}{=} t^{1+\alpha/2} T_{\alpha,B}(t=1)$. The simplest case is $\alpha = 1$, which corresponds to the area under a Brownian excursion: this can be easily seen by permuting the pre-minimum and the post-minimum part of a Brownian bridge, i.e. by using Vervaat's construction [36]. Hence $T_{\alpha=1,B}(t)$ is distributed according to the Airy distribution, discussed above [19,20]. For $\alpha = -1/2$, this yields again the maximal time spent by particles to pass through a disordered periodic potential which is a Brownian bridge. Finally we conjecture, following the lines of reasoning of Refs. [30,33], that for $\alpha = -1$, i. e. the equivalent of I in (4) where $x(\tau)$ is replaced by $x_B(\tau)$ and x_{\max} by $x_{\max,B}$, $T_{\alpha=-1,B}(t)$ yields the cost of the optimal algorithm to find the maximum of a random walk bridge (see Appendix C.3). As shown below, this random variable turns out to be related to the maximum of a Brownian excursion (see also Refs. [34,35]).

2 Summary of main results

The goal of this work is to present various tools to study the statistics of such functionals $\mathcal{O}_{\max}(t)$ of the maximum of Brownian motion and its variants. It is useful to summarize the different approaches developed here as well as the main results obtained in the present paper. The first natural observable to compute is the average value of such functionals, $\langle \mathcal{O}_{\max}(t) \rangle$. According to Eq. (2), the average can be obtained by computing the average density of states $\langle \rho(r, t) \rangle$. As we show here, this quantity can be calculated using rather elementary computations, based on the propagator of the Brownian motion with appropriate boundary conditions. In particular, this method allows us to recover in a very simple way the result for c_0 given in Eq. (5) and previously obtained in Refs. [31] and [32] from the analysis of a rather complicated double integral. The expression $\langle \rho(r, t) \rangle$ for the free BM, as well as for the bridge, was recently announced by us in a short Letter [29]. In this paper, in addition to providing details of these computations, we extend our techniques in several directions obtaining many new results. In particular, we show that this method, relying on propagators, can be easily adapted to a variety of other constrained Brownian motions, including the excursion, the meander, as well as the reflected BM and the reflected Brownian bridge. The main characteristics of $\langle \rho(r, t) \rangle$ for these various constrained BM are summarized in Table 2. From it, we obtain in particular from (2) the average of the functional $T_\alpha(t)$ for the free BM, with $\alpha \in]-2, +\infty[$ as

$$\langle T_\alpha(t) \rangle = \frac{(2t)^{1+\frac{\alpha}{2}} (2 - 2^{-\alpha}) \Gamma\left(\frac{1+\alpha}{2}\right)}{(2+\alpha)\sqrt{\pi}}, \quad (14)$$

from which we get in particular the constant c_0 in Eq. (4) as $c_0 = T_{\alpha=-1}(t=1)/2 = \sqrt{8/\pi} \log 2 = 1.1061\dots$ as given in Eq. (5). For $\alpha = 1$, one recovers $\langle T_{\alpha=1}(t=1) \rangle = \sqrt{2/\pi}$ [20] for the average area under a Brownian double meander (see Fig. 8). As a function of α it has an interesting non-monotonic behavior, diverging when $\alpha \rightarrow -2$ as $\langle T_{\alpha}(t=1) \rangle \sim 4/(2+\alpha)$ as well as when $\alpha \rightarrow \infty$ as $\langle T_{\alpha}(t=1) \rangle \sim 4\sqrt{2}(\alpha/e)^{\alpha/2}/\alpha$, exhibiting a minimum for $\alpha \approx 1.148$. Similarly, for the bridge one obtains

$$\langle T_{\alpha,B}(t) \rangle = \frac{t^{1+\alpha/2}}{2^{\alpha/2}} \Gamma\left(1 + \frac{\alpha}{2}\right). \quad (15)$$

In particular, for $\alpha = 1$ it yields back the first moment of the Airy-distribution, $\langle T_{\alpha=1,B}(t=1) \rangle = \sqrt{\pi/8}$ [37,38] while for $\alpha = -1$ it gives the equivalent of c_0 in Eq. (4) for the Bridge, $c_{0,B} = T_{\alpha=-1,B}(t=1)/2 = \sqrt{\pi/2} = 1.25331\dots$. Interestingly, this means that, on average, the cost of Odlyzko's algorithm is higher for the BB than for the free BM. This can be roughly understood through the fact that the DOS close to the maximum is slightly higher for the BB, which is pinned to the origin on both sides of the time interval, than for the BM which is free on one side. As a function of α , $\langle T_{\alpha,B}(t) \rangle$ is also non-monotonic diverging when $\alpha \rightarrow -2$ as $\langle T_{\alpha,B}(t=1) \rangle \sim 4/(2+\alpha)$ as well as when $\alpha \rightarrow \infty$ as $\langle T_{\alpha,B}(t=1) \rangle \sim 2^{-\alpha}(\alpha/e)^{\alpha/2}\sqrt{\pi\alpha}$, with a minimum for $\alpha \approx 2.960$.

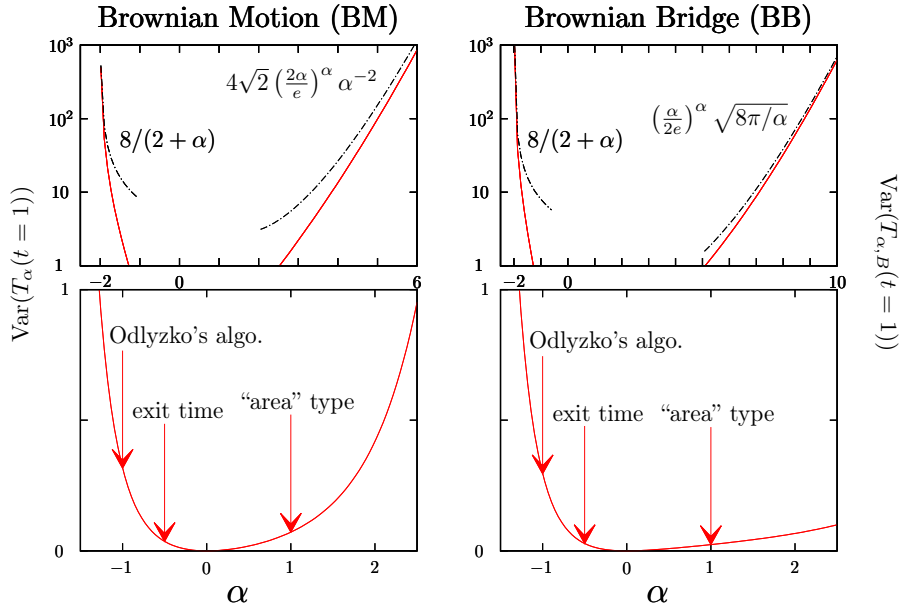


Fig. 3 **Left panel (bottom):** Plot of the variance $\text{Var}(T_{\alpha}(t=1)) = \langle T_{\alpha}^2(t=1) \rangle - \langle T_{\alpha}(t=1) \rangle^2$ as a function of α for the BM, as given in Eq. (16). **Left panel (top):** Focus on the asymptotic behaviors of $\text{Var}(T_{\alpha}(t=1))$ for $\alpha \rightarrow -2$ and $\alpha \rightarrow \infty$. The dashed lines indicate the asymptotic behaviors given in the text below Eq. (16). **Right panel (bottom):** same quantity for the Brownian bridge, $\text{Var}(T_{\alpha,B}(t=1))$, as given in Eq. (17). **Right panel (top):** Focus on the asymptotic behaviors of $\text{Var}(T_{\alpha,B}(t=1))$ for $\alpha \rightarrow -2$ and $\alpha \rightarrow \infty$. The dashed lines indicate the asymptotic behaviors given in the text below Eq. (17). The functionals $T_{\alpha}(t)$ and $T_{\alpha,B}(t)$ interpolate between different interesting observables: for $\alpha = -1$ they describe the cost of Odlyzko's optimal algorithm to find the maximum of a discrete random walk, for $\alpha = -1/2$, they describe the largest exit time of a particle through a Brownian potential and for $\alpha = 1$, $T_{\alpha=1}(t)$ and $T_{\alpha=1,B}(t)$ are random variables of the “area” type. For the bridge (right panel), this corresponds to the area under the BE (Airy random variable).

If one is interested not only in the average of the functional $T_{\alpha}(t)$ but also in higher cumulants or even in its full distribution, the knowledge of the average DOS is obviously not enough and these computations based on propagators become cumbersome. Instead, we present here a method, based

on path-integral which allows us to compute its Laplace transform, namely $\langle e^{-\lambda \mathcal{O}_{\max}(t)} \rangle$, from which the moments of arbitrary order can be obtained by differentiation with respect to (wrt) λ . In some cases, this can also allow us to compute the full distribution of the functional. We first illustrate this method on the DOS itself, for which $V(x) = \delta(x - r)$. The main results in this case were recently announced in Ref. [29], without any details, which we thus provide here. Then we use this general formalism to study the special functionals $T_\alpha(t)$ (12) as well as $T_{\alpha,B}(t)$ in (13) for the BB. In particular, we show how this method allows us to compute the second moment, and eventually the variance $\text{Var}(T_\alpha(t)) = \langle T_\alpha^2(t) \rangle - \langle T_\alpha(t) \rangle^2$ for arbitrary $\alpha \in]-2, +\infty[$ under the form:

$$\text{Var}(T_\alpha(t)) = \frac{t^{2+\alpha}}{(2^{3\alpha}(\alpha+2)^2)} \left(\frac{(2^\alpha - 1)(2^{\alpha+1} - 1)\Gamma(\alpha+3)}{(\alpha+1)^2} - \frac{2^{2\alpha+2}(2^{\alpha+1} - 1)^2 \Gamma(\frac{\alpha+1}{2})^2}{\pi} \right. \\ \left. + \frac{(\alpha+2)\Gamma(2\alpha+2)(2^{2\alpha+2} - 2^{\alpha+1}(\alpha+1)B_{1/2}(\alpha+2, -2(\alpha+1)) - 1)}{(\alpha+1)^2 \Gamma(\alpha+1)} \right), \quad (16)$$

where $B_z(a, b) = \int_0^z t^{a-1}(1-t)^{b-1} dt$ is the incomplete beta function. Formula (16) yields in particular $\text{Var}(T_{\alpha=-1}(t=1)) = \frac{\pi^2}{3} + (4 - 32/\pi) \log(2)^2$, as obtained previously in Ref. [33], using probabilistic methods. For $\alpha = 1$, Eq. (16) yields $\text{Var}(T_{\alpha=1}(t=1)) = 17/24 - 2/\pi$ [20]. Interestingly, as a function of α , it has a non monotonic behavior. It is diverging when $\alpha \rightarrow -2$ as $\text{Var}(T_\alpha(t=1)) \sim 8/(\alpha+2)$ as well as when $\alpha \rightarrow \infty$ as $4\sqrt{2}(2\alpha/e)^\alpha \alpha^{-2}$ exhibiting a single minimum for $\alpha = 0$. In Fig. 3 (left panel), we show a plot of $\text{Var}(T_\alpha(t=1))$ as a function of α .

Similarly, one can also compute the second moment in the case of the bridge, yielding the variance $\text{Var}(T_{\alpha,B}(t)) = \langle T_{\alpha,B}^2(t) \rangle - \langle T_{\alpha,B}(t) \rangle^2$:

$$\text{Var}(T_{\alpha,B}(t)) = t^{\alpha+2} \left[\frac{\sqrt{\pi}(\Gamma(2\alpha+3) - \Gamma(\alpha+2)^2)}{(\alpha+1)^2 \Gamma(\alpha + \frac{3}{2}) 2^{3\alpha+1}} - \frac{\Gamma(\frac{\alpha}{2} + 1)^2}{2^\alpha} \right]. \quad (17)$$

In particular, we can check that $\text{Var}(T_{\alpha=1,B}(t=1)) = 5/12 - \pi/8$ which is the variance of the Airy distribution [19, 20, 37, 38], while for $\alpha = -1$, one has $\text{Var}(T_{\alpha=-1,B}(t=1)) = 2\pi^2/3 - 2\pi$. As a function of α it has also a non-monotonic behavior, diverging when $\alpha \rightarrow -2$ as $\text{Var}(T_{\alpha,B}(t=1)) \sim 8/(\alpha+2)$ and when $\alpha \rightarrow \infty$ as $(\frac{\alpha}{2e})^\alpha \sqrt{8\pi/\alpha}$, exhibiting a minimum for $\alpha = 0$. In Fig. 3 (right panel), we show a plot of $\text{Var}(T_{\alpha,B}(t=1))$ as a function of α .

Finally, in the special case $\alpha = -1$, which corresponds to the cost of the Odlyzko's algorithm, we are able to compute exactly the moments of arbitrary order, both for the free BM, $\langle T_{\alpha=-1}^k(t=1) \rangle$ and for the bridge $\langle T_{\alpha=-1,B}^k(t=1) \rangle$. One obtains indeed,

$$\langle T_{\alpha=-1}^k(t=1) \rangle = \Gamma\left(\frac{k+1}{2}\right) \frac{2^{\frac{k}{2}+2}}{\sqrt{\pi}} \sum_{m=0}^k \tilde{\zeta}(m) \tilde{\zeta}(k-m), \quad \tilde{\zeta}(m) = (1 - 2^{1-m})\zeta(m) = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n^m}, \quad (18)$$

recovering, using a completely different method, the result of Chassaing, Marckert and Yor [33]. For $k = 1$ and $k = 2$, this formula (18) yields back the above results for the first moment (14) and for the variance (16).

For the bridge, we obtain the result, for any real k

$$\langle T_{\alpha=-1,B}^k(t=1) \rangle = -2^{1+\frac{k}{2}} \pi^{k-\frac{1}{2}} k \Gamma\left(\frac{3}{2} - \frac{k}{2}\right) \zeta(1-k). \quad (19)$$

Of course for $k = 1$ and $k = 2$, this formula (19) yields back the aforementioned results for the first moment (15) and the variance (17). In Ref. [33], the authors obtained the full probability distribution function (PDF) of $T_{\alpha=-1}(t=1)$ in terms of a convolution of two theta-functions. Here, we obtain the full PDF $p_B(s)$ of $T_{\alpha=-1,B}(t=1)$ in the case of the BB, as:

$$p_B(s) = \frac{d}{ds} \left(2 \sum_{m=0}^{\infty} (1 - m^2 s^2) e^{-\frac{m^2 s^2}{2}} \right) = \frac{d}{ds} \left(\frac{8\sqrt{2}\pi^{5/2}}{s^3} \sum_{m=1}^{\infty} m^2 e^{-\frac{2m^2 \pi^2}{s^2}} \right), \quad (20)$$

where the two formulas are related to each other via the Poisson summation formula. Interestingly, in Eq. (20), one actually recognizes the PDF of the maximum of a Brownian excursion on the unit time interval, $x_{\max,E} = \max_{0 \leq \tau \leq 1} x_E(\tau)$. One has indeed

$$p_B(s) = \frac{d}{ds} \Pr \left(x_{\max,E} \leq \frac{s}{2} \right). \quad (21)$$

As explained below, one can show that $T_{\alpha=-1,B}(1) \stackrel{\text{law}}{=} \int_0^1 d\tau / x_E(\tau)$. Hence, Eqs. (20) and (21) is a manifestation of a non-trivial identity in law for the Brownian excursion $x_E(\tau)$ [34,35]:

$$\int_0^1 \frac{d\tau}{x_E(\tau)} \stackrel{\text{law}}{=} 2 \max_{0 \leq \tau \leq 1} x_E(\tau). \quad (22)$$

Therefore, our result for $T_{\alpha=-1,B}(t)$ in Eq. (20) provides a simple derivation of this non-trivial identity (22), which was proved in Refs. [34,35] using rather involved probabilistic tools.

The paper is organized as follows. In section 3, we focus on a method based on "counting paths", using propagators of BM. We first illustrate this approach to compute the average DOS for BM $\langle \rho(r, t) \rangle$ in subsection 3.1 and then extend it to the Brownian bridge in section 3.2 as well as to other constrained BMs in 3.3,3.4 and Appendix B. In section 3.4, we present a comparison between our exact results with numerical simulations, the details of which are given in Appendix E. In section 4, we develop a general path integral approach to compute functionals of the maximum of Brownian motion and Brownian bridge. Within this framework, which is presented in detail in section 4.1, we study more specifically the full statistics DOS in section 4.2.1, and the family of functional $T_\alpha(t)$ (cf. Eq. 12) in section 4.2.2. We also analyze more precisely the case $\alpha = -1$ in section 4.2.3 which is relevant to study Odlyzko's algorithm. Some technical details, including a description of the main ideas behind Odlyzko's algorithm, have been left in Appendices.

3 The method of propagators

In this section, we present a rather simple method, based on "counting paths", to compute the average DOS $\langle \rho(r, t) \rangle$. In the cases of BB and BE (which turn out to be identical as we show below), this method allows us also to compute higher moments $\langle \rho_B^k(r, t) \rangle$ with k an arbitrary integer. We illustrate the method in detail for the case of free BM and then apply it to various constrained BM: the Brownian bridge and Brownian excursion in section 3.2 and the Brownian meander in section 3.3. In Appendix B we use the method, for completeness, for the reflected BM and for the reflected BB.

3.1 The average DOS for free Brownian motion

To compute $\langle \rho(r, t) \rangle$, we simply average Eq. (1) over the trajectories of the BM and write it as

$$\langle \rho(r, t) \rangle = \int_0^t \langle \delta(x_{\max} - x(\tau) - r) \rangle d\tau. \quad (23)$$

In Eq. (23) we recognize that the integrand $\langle \delta(x_{\max} - x(\tau) - r) \rangle$ has a simple probabilistic interpretation. Indeed, one has

$$\langle \delta(x_{\max} - x(\tau) - r) \rangle d\tau = \Pr [x(\tau) \in [x_{\max} - r - d\tau, x_{\max} - r]] . \quad (24)$$

The idea to compute the PDF $\langle \delta(x_{\max} - x(\tau) - r) \rangle$ in (24) is to evaluate the "number" of paths that reach their maximum M at time t_{\max} , pass through $M - r$ at time τ , and end at $x_F \leq M$ at time t . The total number which we want to compute is then obtained by integrating over x_F, M and t_{\max} (see Fig. 4). In each of the three time intervals delimited by τ, t_{\max} and t , the BM is constrained to stay below M . This number of paths can thus be computed from the propagator $G_M(\alpha|\beta, t)$ of the BM, starting at time 0 at $x = \alpha < M$ and arriving at time t at $x = \beta < M$ and staying below M during the whole time interval $[0, t]$. It can be easily computed using, for instance, the method of images:

$$G_M(\alpha|\beta, t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{(\beta-\alpha)^2}{2t}} - e^{-\frac{(2M-\beta-\alpha)^2}{2t}} \right). \quad (25)$$

For future purpose, it is useful to compute its Laplace transform (LT) wrt t , $\tilde{G}_M(\alpha|\beta, s)$. It is given by

$$\tilde{G}_M(\alpha|\beta, s) = \int_0^\infty e^{-st} G_M(\alpha|\beta, t) dt = \frac{1}{\sqrt{2s}} \left(e^{-\sqrt{2s}|\beta-\alpha|} - e^{-\sqrt{2s}(2M-\beta-\alpha)} \right). \quad (26)$$

When dividing the time interval $[0, t]$ into three parts as in Fig. 4, two cases may arise: $\tau < t_{\max}$ or

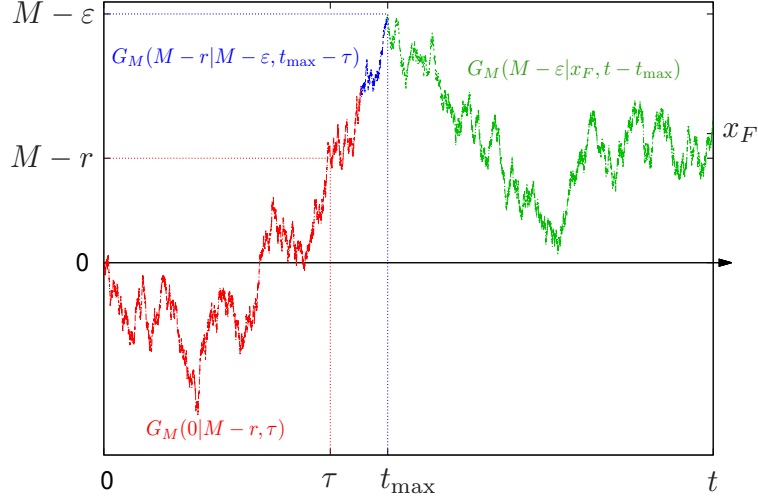


Fig. 4 (Color online) Illustration of the method of propagators. The BM starts at 0 and visits the point $M-r$ at time τ , passes through $M-\varepsilon$ at time t_{\max} and ends at x_F at time t .

$\tau > t_{\max}$. One can show that these two configurations give rise to the same contributions to $\langle \rho(r, t) \rangle$: this can be seen by making a global shift $x(\tau) \rightarrow x(\tau) - x_F/2$ (for each realization of BM) and using the time reversal symmetry. Note also that, when dealing with BM which is continuous both in space and time, one can not impose simultaneously $x(t_{\max}) = M$ and $x(t) < M$ right before and after t_{\max} . To circumvent this difficulty, one imposes instead $x(t_{\max}) = M - \varepsilon$, and take the limit $\varepsilon \rightarrow 0$ at the end of the calculation. Finally, $\langle \rho(r, t) \rangle$ is given by the ratio of paths which have the analyzed properties to the normalization constant $Z(\varepsilon)$ which count the same paths without the condition to pass through $M-r$ at time τ

$$\langle \rho(r, t) \rangle = \lim_{\varepsilon \rightarrow 0} \frac{2}{Z(\varepsilon)} \int_0^\infty dM \int_0^t dt_{\max} \int_{-\infty}^M dx_F \int_0^{t_{\max}} d\tau \quad G_M(0|M-r, \tau) G_M(M-r|M-\varepsilon, t_{\max} - \tau) \\ \times G_M(M-\varepsilon|x_F, t - t_{\max}), \quad (27)$$

where we have used the Markov property of BM and where the factor of 2 comes from the two aforementioned equivalent situations corresponding $\tau < t_{\max}$ and $\tau > t_{\max}$. In (27) the normalization constant $Z(\varepsilon)$ is given by

$$Z(\varepsilon) = \int_0^\infty dM \int_0^t dt_{\max} \int_{-\infty}^M dx_F G_M(0|M-\varepsilon, t_{\max}) G_M(M-\varepsilon|x_F, t - t_{\max}). \quad (28)$$

The normalization is easily computed as $Z(\varepsilon) \sim 2\varepsilon^2$ when $\varepsilon \rightarrow 0$. In (27), we recognize a convolution structure. Taking the LT wrt to t , we find

$$\int_0^\infty dt e^{-st} \langle \rho(r, t) \rangle = \lim_{\varepsilon \rightarrow 0} \frac{2}{Z(\varepsilon)} \int_0^\infty dM \int_{-\infty}^M dx_F \tilde{G}_M(0|M-r, s) \tilde{G}_M(M-r|M-\varepsilon, s) \tilde{G}_M(M-\varepsilon|x_F, s). \quad (29)$$

Using (26) and performing a small ε expansion, one obtains straightforwardly

$$\begin{aligned} \int_0^\infty dt e^{-st} \langle \rho(r, t) \rangle &= 8 \left[\int_0^r dM \frac{e^{-\sqrt{2s}r}}{\sqrt{2s}} \sinh(\sqrt{2s}M) e^{-\sqrt{2s}r} \int_{-\infty}^M dx_F e^{-\sqrt{2s}(M-x_F)} \right. \\ &\quad \left. + \int_r^\infty dM \frac{e^{-\sqrt{2s}M}}{\sqrt{2s}} \sinh(\sqrt{2s}r) e^{-\sqrt{2s}r} \int_{-\infty}^M dx_F e^{-\sqrt{2s}(M-x_F)} \right]. \end{aligned} \quad (30)$$

Performing the remaining integrals over x_F and M we obtain

$$\int_0^\infty dt e^{-st} \langle \rho(r, t) \rangle = 8 \frac{e^{-\sqrt{2s}r} - e^{-2\sqrt{2s}r}}{(2s)^{3/2}}. \quad (31)$$

By inverting the above LT (31) we finally obtain

$$\langle \rho(r, t) \rangle = \sqrt{t} \bar{\rho} \left(\frac{r}{\sqrt{t}}, 1 \right), \quad \bar{\rho}(r, t=1) = 8 \left(\Phi^{(2)}(r) - \Phi^{(2)}(2r) \right), \quad \Phi^{(2)}(r) = \frac{e^{-\frac{r^2}{2}}}{\sqrt{2\pi}} - \frac{r}{2} \operatorname{erfc} \left(\frac{r}{\sqrt{2}} \right), \quad (32)$$

where $\Phi^{(2)}$ belongs to a useful hierarchy of functions, $\Phi^{(k)}$, as explained in Appendix A. From the average value of the DOS, one can compute the average value of any functional of the maximum, according to (2). In particular, for the special case $V(x) = x^\alpha$, one obtains from (2) and (32):

$$\begin{aligned} \langle T_\alpha(t) \rangle &= \left\langle \int_0^t [x_{\max} - x(\tau)]^\alpha d\tau \right\rangle = \int_0^\infty \langle \rho(r, t) \rangle r^\alpha dr \\ &= 8t^{1+\alpha/2} \int_0^\infty r^\alpha \left(\Phi^{(2)}(r) - \Phi^{(2)}(2r) \right) dr = \frac{(2t)^{1+\frac{\alpha}{2}} (2 - 2^{-\alpha}) \Gamma(\frac{1+\alpha}{2})}{(2+\alpha)\sqrt{\pi}}, \end{aligned} \quad (33)$$

as announced in the introduction in Eq. (14). In Fig. 5 (left panel), we show a plot of $\langle T_\alpha(t) \rangle$ as a function of α . Specifying this formula (33) to the case $\alpha = -1$, one obtains $c_0 = T_{\alpha=-1}(t=1)/2 = \sqrt{\frac{8}{\pi}} \log 2$ as announced in (5), recovering in a rather simple way the result of Refs. [31,32].

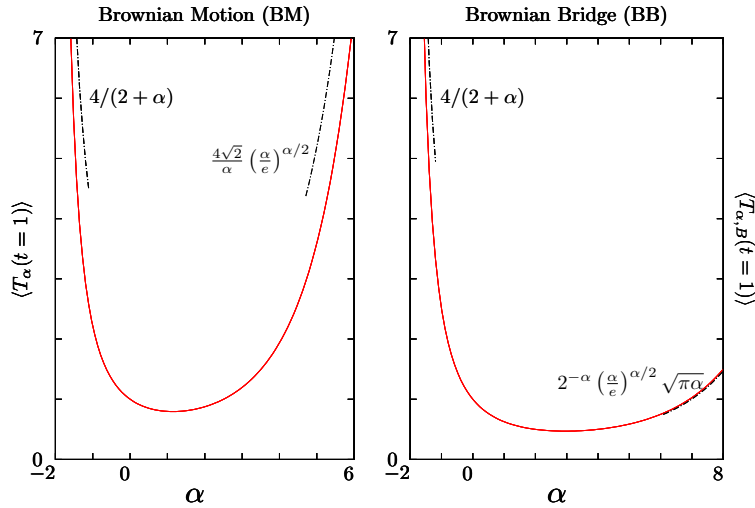


Fig. 5 Left panel: Plot of $\langle T_\alpha(t=1) \rangle$, as a function of α , for the BM, as given in Eq. (33). **Right panel:** Plot of $\langle T_{\alpha,B}(t=1) \rangle$, as a function of α , for the BB, as given in Eq. (44). In both panels, the dashed black lines indicate the asymptotic behaviors discussed in the introduction. In particular, for $\alpha = -1$ for the BM (corresponding to Odlyzko's algorithm), we recover $\langle T_{\alpha=-1}(t=1) \rangle = 4\sqrt{2/\pi} \log 2$ [31,32]. For $\alpha = 1$ for the BB (corresponding to the Airy distribution), we recover $\langle T_{\alpha=1,B}(t=1) \rangle = \sqrt{\pi/8}$ [19,20].

This method can in principle be adapted to compute higher moments of the DOS, $\langle \rho^k(r, t) \rangle$, but such computations involve a rather cumbersome combinatorial analysis, for $k > 1$. A more powerful approach amounts instead to compute the generating function of $\rho(r, t)$ [29] – as shown in section 4. As we show now, this method can also be easily extended to compute the average DOS for various constrained BM.

3.2 The average DOS and its higher moments for Brownian bridges and Brownian excursions

For Brownian bridges, as well as Brownian excursions, the method based on propagators of constrained BM, allows us to compute arbitrary moments of the DOS, $\langle \rho_B^k(r, t) \rangle$. An expression for these moments can be written from the definition [see Eq. (1) where $x(\tau)$ is replaced by the BB $x_B(\tau)$] as

$$\langle \rho_B^k(r, t) \rangle = \int_0^t \dots \int_0^t \left\langle \prod_{i=1}^k dt_i \delta(x_{\max, B} - x_B(t_i) - r) \right\rangle, \quad (34)$$

such that in (34) the BB visits k times the point $M - r$ at successive times t_1, t_2, \dots, t_k where M is the value of the maximum on $[0, t]$. To compute this quantity (34) for the BB, it is useful to invoke the Vervaat's construction to relate the DOS of the Brownian bridge to the *local time* of the Brownian excursion (see Fig. 6).

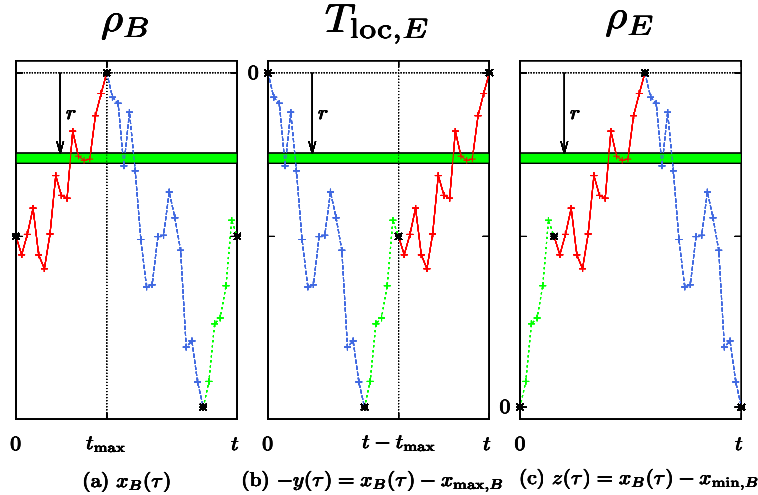


Fig. 6 (Color online) Illustration of the identities in law given in Eq. (35) in the text relating the DOS near the maximum for the BB to the local time, on the one hand, and to the DOS, on the other hand, for the BE. **(a)** A typical trajectory of a BB, $x_B(\tau)$. The process spends a time $\rho_B(r, t)dr$ in the green stripe $[x_{\max, B} - r - dr, x_{\max, B} - r]$. **(b)** The process $-y(\tau) = x_{\max, B} - x_B(\tau)$ after the Vervaat's transformation explained in the text is an excursion. The time spent in the green stripe is now given by $T_{\text{loc}, E}(r, t)dr$ for this new process $y(\tau)$. **(c)** The process $z(\tau) = x_B(\tau) - x_{\min, B}$ after the transformation explained in the text is also an excursion. The time spent in the green stripe is now given by $\rho_E(r, t)dr$ for this new process.

The Vervaat's construction works as follows [36]: we denote by t_{\max} the time at which the BB reaches its maximum on $[0, t]$ (see Fig. 6 (a)) and break the time interval $[0, t]$ into two parts, $[0, t_{\max}]$ and $[t_{\max}, t]$ as illustrated in Fig. 6 (a). Then we permute the two associated portions of the path, the continuity of the path being guaranteed by $x_B(t) = x_B(0) = 0$ for the BB. We can further transform the path by considering $y(\tau) = x_{\max, B} - x_B(\tau)$ and take finally the origin of times at t_{\max} : this yields the configuration shown in Fig. 6 (b). Furthermore, if we initially break the time interval into $[0, t_{\min}]$ and $[t_{\min}, t]$, where t_{\min} denotes the time at which the BB reaches its minimum on $[0, t]$, we permute

the two associated portions of the path and finally take the origin of times at t_{\min} , we obtain the configuration shown in Fig. 6 (c).

These transformations show that $\rho_E(r, t)$ for an excursion is identical in law to the local time in r for an excursion, $T_{\text{loc}, E}(r, t)$ and hence also identical in law to the DOS of the BB. Hence the Vervaat's construction allows to us write the following identities:

$$\rho_B(r, t) \stackrel{\text{law}}{=} \rho_E(r, t) \stackrel{\text{law}}{=} T_{\text{loc}, E}(r, t) = \int_0^t \delta(x_E(\tau) - r) d\tau, \quad (35)$$

where we emphasize that $x_E(\tau)$ is a Brownian excursion. Therefore, the moments of the DOS for the bridge or the excursion can be computed as

$$\langle \rho_B^k(r, t) \rangle = \langle \rho_E^k(r, t) \rangle = \langle \prod_{i=1}^k \int_0^t dt_i \delta(x_E(t_i) - r) \rangle. \quad (36)$$

As done before for the free BM in Eq. (27), the expression in (36) can be computed using the propagator of the excursion, i.e. the propagator $G^+(\alpha|\beta, t)$ of a free BM starting at $\alpha > 0$ at time $t = 0$ and reaching $\beta > 0$ at time t and staying positive over the whole interval $[0, t]$. This propagator $G^+(\alpha|\beta, t)$ can also be computed simply by the method of images, yielding:

$$G^+(\alpha|\beta, t) = \frac{1}{\sqrt{2\pi t}} \left(e^{-\frac{(\alpha-\beta)^2}{2t}} - e^{-\frac{(\alpha+\beta)^2}{2t}} \right), \quad (37)$$

together with its LT transform wrt t , which will be useful in the following:

$$\tilde{G}^+(\alpha|\beta, s) = \frac{1}{\sqrt{2s}} \left(e^{-\sqrt{2s}|\alpha-\beta|} - e^{-\sqrt{2s}(\beta+\alpha)} \right). \quad (38)$$

Since we are studying here a Brownian excursion, we need, as before, to introduce a regulator such that $x_E(0) = x_E(t) = \varepsilon > 0$ and take the limit $\varepsilon \rightarrow 0$ of a suitably defined quantity, as done in Eq. (27) – since we can not impose simultaneously $x_E(0) = 0$ and $x_E(0^+) > 0$. One has here:

$$\langle \rho_B^k(r, t) \rangle = \langle \rho_E^k(r, t) \rangle = \lim_{\varepsilon \rightarrow 0} \frac{k!}{Z_B(\varepsilon)} \int_{0 < t_1 < t_2 < \dots < t_k < t} dt_1 \dots dt_k G^+(\varepsilon|r, t_1) G^+(r|r, t_2 - t_1) \dots G^+(r|\varepsilon, t - t_k), \quad (39)$$

where the combinatorial factor $k!$ comes from the different permutations of the intermediate times t_1, \dots, t_k where the Brownian excursion $x_E(\tau)$ reaches the value r . The denominator $Z_B(\varepsilon)$ is given by

$$Z_B(\varepsilon) = G^+(\varepsilon|\varepsilon, t) \sim \varepsilon^2 \sqrt{\frac{2}{\pi}} t^{-3/2}, \quad \varepsilon \rightarrow 0. \quad (40)$$

To compute the multiple integral over the times t_1, \dots, t_k in the right hand side of Eq. (39), we recognize, as before for the free BM, a convolution structure. Hence its LT wrt t is given by:

$$\begin{aligned} \int_0^\infty dt e^{-st} \int_{0 < t_1 < t_2 < \dots < t_k < t} dt_1 \dots dt_k G^+(\varepsilon|r, t_1) G^+(r|r, t_2 - t_1) \dots G^+(r|\varepsilon, t - t_k) &= \tilde{G}^+(\varepsilon|r, s)^2 \tilde{G}^+(r|r, s)^{k-1} \\ &\sim 4\varepsilon^2 e^{-2\sqrt{2sr}} \frac{(1 - e^{-2\sqrt{2sr}})^{k-1}}{(2s)^{(k-1)/2}}, \quad \varepsilon \rightarrow 0, \end{aligned} \quad (41)$$

where we have used the explicit expression of the LT of the propagator in (38). Using the above expression (41) together with (40), one obtains finally after Laplace inversion:

$$\langle \rho_B^k(r, t = 1) \rangle = \langle \rho_E^k(r, t = 1) \rangle = \langle T_{\text{loc}, E}^k(r, t = 1) \rangle = 2\sqrt{2\pi} k! \sum_{l=1}^k (-1)^{l+1} \binom{k-1}{l-1} \Phi^{(k-2)}(2rl), \quad (42)$$

with the convention $\Phi^{(-1)} = -d\Phi^{(0)}/dr$ and where the $\Phi^{(j)}$'s are defined in Appendix A, see Eq. (142). Thus we recover here (42) the result obtained by Takacs [39] (see his Eq. (38)) by a probabilistic

method (note the correspondence between the functions J_k in [39] and the functions Φ^k : $J_k(r) = r^{1+k} \int_1^\infty dx e^{-(xr)^2/2} (x-1)^k = \sqrt{2\pi k!} \Phi^{(k+1)}(r)$).

For $k=1$, one finds the mean DOS for the BB on the unit time interval, $\langle \rho_B(r, t=1) \rangle = 4re^{-2r^2}$, as found in Ref. [39]. Note that it coincides in this case with the PDF of the maximum of a BB, which is a generic property for periodic signals such that $x(t) = x(0)$ [40].

$$\begin{aligned} \langle \rho_B(r, t=1) \rangle &= \langle \rho_E(r, t=1) \rangle = \int_0^1 d\tau \langle \delta(x_{\max, B} - x_B(\tau) - r) \rangle \\ &= \langle \delta[x_{\max, B} - r] \rangle \\ &= 4re^{-2r^2}, \end{aligned} \quad (43)$$

where we have used the periodicity of the bridge and the possibility of adding to it an arbitrary constant. From the average DOS, we can compute, from (2) the average value of any functional of the maximum of a BB. In particular, for the interesting family of functionals $V(x) = x^\alpha$, one obtains

$$\langle T_{\alpha, B}(t) \rangle = \left\langle \int_0^t (x_{\max, B} - x_B(\tau))^\alpha d\tau \right\rangle = \int_0^\infty \langle \rho_B(r, t) \rangle r^\alpha dr \quad (44)$$

$$= t^{1+\frac{\alpha}{2}} \frac{\Gamma(1+\frac{\alpha}{2})}{2^{\alpha/2}}, \quad (45)$$

as announced in the introduction in Eq. (15). In Fig. 5 (right panel), we show a plot of $\langle T_{\alpha, B}(t) \rangle$ as a function of α .

3.3 The average DOS for the Brownian meander

Using the same method based on propagators (27), we can also compute the average DOS for the Brownian meander (see Fig. 1 (d)). To this purpose, we need to know the propagator of a Brownian particle confined in a given interval $[0, M]$ with absorbing boundary conditions both in $x=0$ and $x=M$. Denoting by $G_M^+(\alpha|\beta, t)$ the propagator of such a particle starting at α and ending, at time t , at β , one has

$$G_M^+(\alpha|\beta, t) = \sum_{n=1}^{\infty} \frac{2}{M} \sin\left(\frac{\pi n}{M}\alpha\right) \sin\left(\frac{\pi n}{M}\beta\right) e^{-\frac{\pi^2}{2M^2}n^2 t}. \quad (46)$$

Its LT wrt t , $\tilde{G}_M^+(\alpha|\beta, s)$ reads

$$\tilde{G}_M^+(\alpha|\beta, s) = \frac{2 \sinh[\sqrt{2s}(M - \max(\alpha, \beta))] \sinh[\sqrt{2s} \min(\alpha, \beta)]}{\sqrt{2s} \sinh(\sqrt{2s}M)}, \quad (47)$$

where we have used the identity

$$\sum_{k=1}^{\infty} \frac{\cos(kx)}{a^2 + k^2} = \frac{1}{2a} \left(\frac{\pi \cosh((\pi - x)a)}{\sinh(\pi a)} - \frac{1}{a} \right). \quad (48)$$

As done before in (27), we introduce the two times t_{\max} and τ such that $x(t_{\max}) = M - \epsilon$ and $x(\tau) = M - r$. These two times break the interval into three sub-intervals (see Fig. 4). As shown in Fig. 4, two cases may arise: $0 < \tau < t_{\max}$ or $t > \tau > t_{\max}$. In this case, for the Brownian meander (BMe), these two configurations do not give the same contributions to the average DOS $\langle \rho_{Me}(r, t) \rangle$. Using the same type of regularization procedure as used before (see Eq. (27)), one has

$$\begin{aligned} \langle \rho_{Me}(r, t) \rangle &= \lim_{\epsilon \rightarrow 0} \frac{1}{Z_{Me}(\epsilon)} \int_r^\infty dM \int_0^t dt_{\max} \int_0^{M-\epsilon} dx_F \\ &\quad \left(\int_0^{t_{\max}} d\tau G_M^+(\epsilon|M-r, \tau) G_M^+(M-r|M-\epsilon, t_{\max}-\tau) G_M^+(M-\epsilon|x_F, t-t_{\max}) \right. \\ &\quad \left. + \int_{t_{\max}}^t d\tau G_M^+(\epsilon|M-\epsilon, t_{\max}) G_M^+(M-\epsilon|M-r, \tau-t_{\max}) G_M^+(M-r|x_F, t-\tau) \right), \end{aligned} \quad (49)$$

	$r \rightarrow 0$	$r \rightarrow \infty$
Brownian (BM)	$4r$	$4\sqrt{\frac{2}{\pi}} \frac{1}{r^2} e^{-\frac{r^2}{2}}$
Bridge/Excursion (BB/BE)	$4r$	$4re^{-2r^2}$
Meander (BMe)	$4r$	$\frac{4}{3} \frac{1}{r} e^{-\frac{r^2}{2}}$
Reflected Brownian	$4r$	$\frac{8}{3} \sqrt{\frac{2}{\pi}} \frac{1}{r^2} e^{-\frac{r^2}{2}}$
Reflected Bridge	$4r$	$3 \frac{1}{r} e^{-2r^2}$

Table 1 Asymptotic behaviors of average DOS for various constrained Brownian motions, both for small and large argument.

where we have used the Markov property of BM and where the normalization constant $Z_{Me}(\varepsilon)$ is given by

$$Z_{Me}(\varepsilon) = \int_0^\infty dM \int_0^t dt_{\max} \int_0^{M-\varepsilon} dx_F G_M^+(0|M-\varepsilon, t_{\max}) G_M^+(M-\varepsilon|x_F, t-t_{\max})$$

$$\sim 2\varepsilon^3 \sqrt{\frac{2}{\pi t}}, \quad \varepsilon \rightarrow 0. \quad (50)$$

To compute the numerator in (49), we take advantage of its convolution structure and perform its LT wrt t , using the expression of the LT of the propagator (47). After some manipulations, we obtain finally

$$\langle \rho_{Me}(r, t=1) \rangle = 2\sqrt{2\pi} \left(\sum_{n=1}^{\infty} \frac{4n(-1)^n}{2n^2 + 3(-1)^n - 5} \Phi^{(1)}(nr) - \Phi^{(1)}(2r) \right), \quad (51)$$

where $\Phi^{(1)}(x) = \text{erfc}(x)/2$ belongs to the family of functions studied in Appendix A.

3.4 Average DOS of constrained Brownian motions: summary and comparison

It is useful to summarize and compare the results for the DOS for the BM and its variants which we have studied here, using a method based on propagators. In Fig. 7, we have plotted (in lines) the results for the DOS for the Brownian motion (32), Brownian bridge and Brownian excursion (42), the Brownian meander (51) as well as for the reflected Brownian motion (151) and reflected Brownian bridge (154), the study of which has been left in Appendix B. We have computed numerically the DOS for these different constrained BMs – which have been simulated using the constructions described in detail in Appendix E. It is useful to remind that $\langle \rho(r, t) \rangle / t = \langle \rho(r, t=1) \rangle$ has the interpretation of a PDF, as $dr \langle \rho(r, t) \rangle / t$ is the probability that the walker lies in the interval $[x_{\max} - r - dr, x_{\max} - r]$.

The main characteristics of these DOS are summarized in Tables 1 and 2. A first interesting feature is that the small r behavior of the DOS is, at leading order, the same for all these constrained BMs, $\rho(r, t=1) \sim 4r$. This indicates that the local vicinity of the maximum of the BM is insensitive to the boundary conditions in space (the presence of a wall at $x=0$ either absorbing or reflecting does not

	$\langle r \rangle$	r_{typ}	numerics	
			$\langle r \rangle$	r_{typ}
Brownian (BM)	$\sqrt{\frac{2}{\pi}} = 0.7979\dots$	0.5145...	0.79 ± 0.01	0.52 ± 0.02
Bridge/Excursion (BB/BE)	$\frac{1}{2} \sqrt{\frac{\pi}{2}} = 0.6267\dots$	0.5	0.62 ± 0.01	0.48 ± 0.02
Meander (BMe)	$\frac{1}{4} \sqrt{\frac{\pi}{2}} (8 \log 2 - 3) = 0.7975\dots$	0.4907...	0.79 ± 0.01	0.48 ± 0.02
Reflected Brownian	$\frac{3\pi-4}{3\sqrt{2\pi}} = 0.7214\dots$	0.5212...	0.72 ± 0.01	0.52 ± 0.02
Reflected Bridge	$\frac{1}{4} \sqrt{\frac{\pi}{2}} (4 \log 2 - 1) = 0.5554\dots$	0.4907...	0.55 ± 0.01	0.48 ± 0.02

Table 2 Main characteristics of the average DOS for various constrained Brownian motions: the average value $\langle r \rangle$ [see Eq. (52)] and the typical value r_{typ} which is the location of the peak of $\langle \rho(r, t=1) \rangle$. The numerical result are obtained by averaging over 4.10^4 realizations of independent RWs of 10^4 steps. We refer the reader to Appendix E for a description of the algorithms which were used here.

affect it) as well as in time (the free BM and the BB, corresponding to periodic boundary conditions in the time direction, share the same local properties). On the other hand, for large argument, the DOS exhibits, in all the cases, a leading order Gaussian decay (see Table 1) but with different rates. In particular, for bridges its decays much faster, $\propto e^{-2r^2}$ than for free processes, which decays as $e^{-r^2/2}$.

This also implies that the mean value

$$\langle r \rangle = \int_0^\infty dr r \langle \rho(r, t=1) \rangle \quad (52)$$

is larger for the bridge process than for the one which is free at the extremity of the time interval. Interestingly enough, although the mean value might differ notably from one process to another (see Table 2), the typical value r_{typ} , which is the location of the peak of $\langle \rho(r, t=1) \rangle$, does not vary too much, with a typical value $r_{\text{typ}} \approx 1/2$ for all the processes (see Table 2 and Fig. 7).

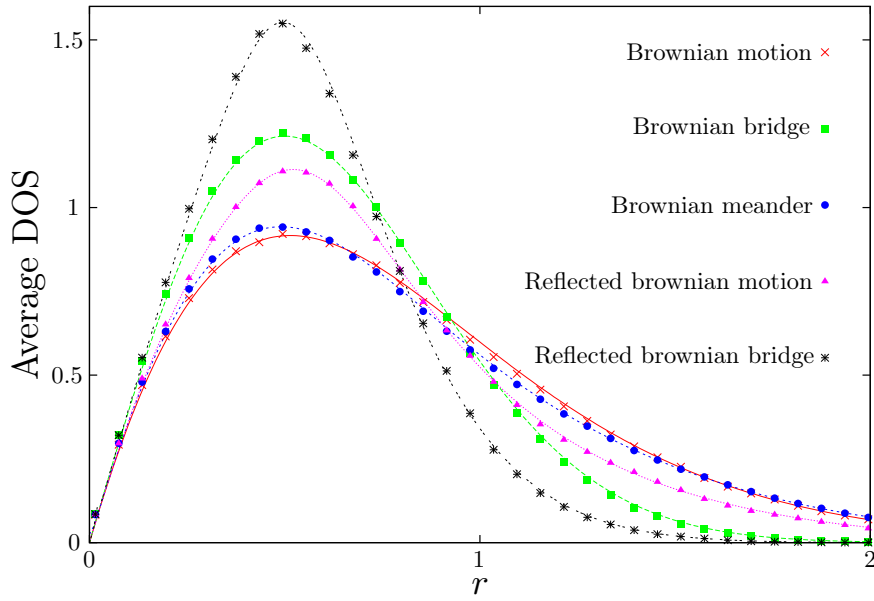


Fig. 7 (Color online) Plot of the average DOS for Brownian motion and its variants. The symbols indicate the results of our numerical simulations (obtained by averaging over 10^4 realizations of independent RWs of 10^4 steps). The lines correspond to our exact analytical results given in Eqs. (32, 42, 51, 151, 154). We refer the reader to Appendix E for a description of the algorithms which were used here.

4 Path integral approach

As mentioned in the introduction, the average DOS is useful to compute the average of any functional of the maximum $\mathcal{O}_{\text{max}}(t) = \int_0^t V(x_{\text{max}} - x(\tau))d\tau$, see Eq. (2). However, if one is interested in higher moments of $\mathcal{O}_{\text{max}}(t)$, the knowledge of the average DOS is not enough. A useful approach to study the statistics of $\mathcal{O}_{\text{max}}(t)$, beyond the first moment, is to compute the Laplace transform of the full PDF of $\mathcal{O}_{\text{max}}(t)$, namely $\langle e^{-\lambda \int_0^t V(x_{\text{max}} - x(\tau))d\tau} \rangle$, from which the moments of $\mathcal{O}_{\text{max}}(t)$ can be obtained by successive derivation wrt λ . In this section, we present a general approach, based on path integral techniques to compute this Laplace transform, for any function $V(x)$. We develop this general framework both for the free BM and then for the Brownian bridge – which turns out to be easier to study, thanks to the Vervaat’s construction (see Fig. 6). We then apply these general methods to study the full statistics of $\rho_{\text{DOS}}(r, t)$ and then to the study of the functionals $T_\alpha(t)$, corresponding to the case

$V(x) = x^\alpha$, for the free BM and for the BB. Then, we focus on the special case $\alpha = -1$, corresponding to Odlyzko's algorithm, which, as we show below, leads to a quantum mechanical problem which is exactly solvable.

4.1 General framework

4.1.1 Free Brownian motion

To study analytically $\langle \exp[-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))] \rangle$, with $\lambda > 0$, for an arbitrary function $V(x)$, we first decompose the time interval $[0, t]$ into two subintervals $[0, t_{\max}]$ and $[t_{\max}, t]$ where t_{\max} is the time at which the maximum is reached. These two intervals $[0, t_{\max}]$ and $[t_{\max}, t]$ are statistically independent (as BM is Markovian), and the PDF of t_{\max} is given by the arcsine law [41],

$$P(t_{\max}) = \frac{1}{\pi \sqrt{t_{\max}(t - t_{\max})}}. \quad (53)$$

The process $y(\tau) = x_{\max} - x(\tau)$ is obviously a BM which stays positive on $[0, t]$. By reversing the time arrow in the interval $[0, t_{\max}]$ and taking t_{\max} as the new origin of time, we see that $y(\tau)$ is built from two independent Brownian meanders (BMe): one of duration t_{\max} and the other (independent) one of duration $t - t_{\max}$ (see Fig. 8). Therefore one has

$$\langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \int_0^t dt_{\max} \varphi(t_{\max}) \varphi(t - t_{\max}) \quad (54)$$

$$\varphi(\tau) = \frac{1}{\sqrt{\pi\tau}} \langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_+, \quad (55)$$

where $\langle \dots \rangle_+$ denotes an average over the trajectories of a BMe $y(\tau)$. In (55) the prefactor $1/\sqrt{\pi\tau}$ comes from the PDF of t_{\max} (53).

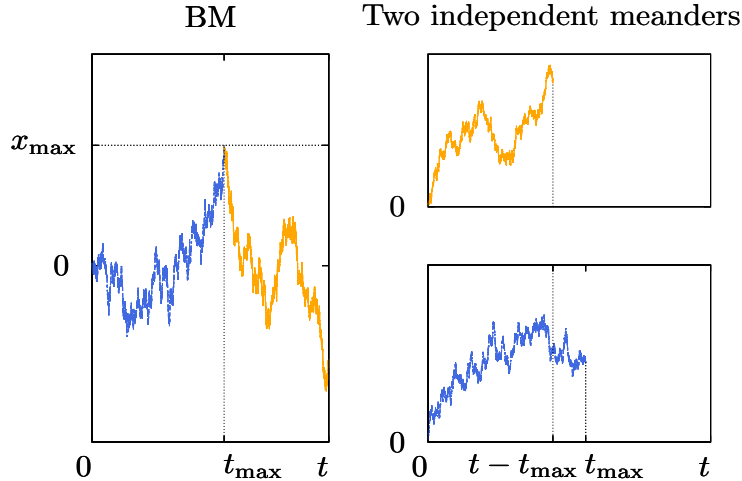


Fig. 8 (Color online) Illustration of the path decomposition used in the path integral method (54): the BM is built from two independent realizations of a Brownian meander, one of duration t_{\max} and the other one of duration $t - t_{\max}$.

On the other hand the convolution structure in (54) suggests to compute its LT wrt t :

$$\int_0^\infty e^{-st} \langle e^{-\lambda \int_0^t V(x_{\max} - x(\tau)) d\tau} \rangle dt = [\tilde{\varphi}(s)]^2, \quad (56)$$

$$\tilde{\varphi}(s) = \int_0^\infty e^{-st} \varphi(t) dt. \quad (57)$$

The next step is to compute $\tilde{\varphi}(s)$ in (57) using a path-integral method. For a Brownian meander, which is continuous both in space and time, this path integral method must be handled with care: as noticed repeatedly in the previous section, one can not impose simultaneously $y(0) = 0$ and $y(0^+) > 0$. This can be circumvented [19] by introducing a cut-off $\varepsilon > 0$ such that $y(0) = \varepsilon$ so that the Feynman-Kac formula reads:

$$\langle e^{-\lambda \int_0^\tau V(y(u)) du} \rangle_+ = \lim_{\varepsilon \rightarrow 0} \frac{\int_0^\infty \langle y_F | e^{-H_\lambda \tau} | \varepsilon \rangle dy_F}{\int_0^\infty \langle y_F | e^{-H_0 \tau} | \varepsilon \rangle dy_F}, \quad (58)$$

with

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dy^2} + \lambda V(y) + V_{\text{wall}}(y), \quad (59)$$

where $V_{\text{wall}}(y)$ is a hard-wall potential, $V_{\text{wall}}(y) = 0$ for $y \geq 0$ and $V_{\text{wall}}(y) = +\infty$ for $y < 0$, which guarantees that the walker stays positive, as it should for a meander. We can easily compute the eigenfunctions of $H_0 = -\frac{1}{2} \frac{d^2}{dy^2} + V_{\text{wall}}(y)$:

$$\psi_k(y) = \sqrt{\frac{2}{\pi}} \sin(ky), \quad (60)$$

and

$$\langle y_F | e^{-H_0 \tau} | \varepsilon \rangle = \frac{2}{\pi} \int_0^\infty dk \sin(ky_F) \sin(k\varepsilon) e^{-\tau \frac{k^2}{2}}. \quad (61)$$

A straightforward computation yields

$$\int_0^\infty dy_F \langle y_F | e^{-H_0 \tau} | \varepsilon \rangle \underset{\varepsilon \rightarrow 0}{\sim} \frac{2\varepsilon}{\sqrt{2\pi\tau}}. \quad (62)$$

We denote by Ψ_n the wave functions of H_λ associated to the energy E_n (we will assume that the spectrum of H_λ is discrete but the computation can easily be extended to a continuum spectrum):

$$\begin{aligned} \langle e^{-\lambda \int_0^\tau du V(y(u))} \rangle_+ &= \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{2\pi\tau}}{2\varepsilon} \int_0^\infty dy_F \sum_{n=0}^\infty \Psi_n^*(y_F) \Psi_n(\varepsilon) e^{-E_n \tau} \\ &= \lim_{\varepsilon \rightarrow 0} \sqrt{\frac{\pi\tau}{2}} \int_0^\infty dy_F \sum_{n=0}^\infty \Psi_n^*(y_F) \frac{\Psi_n(\varepsilon)}{\varepsilon} e^{-E_n \tau} \\ &= \sqrt{\frac{\pi\tau}{2}} \int_0^\infty dy_F \sum_{n=0}^\infty \Psi_n^*(y_F) \Psi_n'(0) e^{-E_n \tau} \\ &= \sqrt{\frac{\pi\tau}{2}} \int_0^\infty dy_F \partial_x G_\tau(0, y_F), \end{aligned} \quad (63)$$

where we have used $\Psi_n(0) = 0$, because of the absorbing wall in 0, and where we denote

$$G_\tau(x, y) = \sum_{n=0}^\infty \Psi_n^*(y) \Psi_n(x) e^{-E_n \tau}. \quad (64)$$

We can use these formulae (63, 64) to calculate the Laplace transform $\tilde{\varphi}(s)$ in (57)

$$\begin{aligned}\tilde{\varphi}(s) &= \int_0^\infty e^{-s\tau} \varphi(\tau) d\tau = \frac{1}{\sqrt{2}} \int_0^\infty d\tau e^{-s\tau} \int_0^\infty dy_F \partial_x G_\tau(0, y_F) = \frac{1}{\sqrt{2}} \int_0^\infty dy_F \sum_{n=0}^\infty \frac{\Psi_n^*(y_F) \Psi_n'(0)}{s + E_n} \\ &= \frac{1}{\sqrt{2}} \int_0^\infty dy_F \partial_x \tilde{G}_s(0, y_F),\end{aligned}\quad (65)$$

where $\tilde{G}_s(x, y) = \sum_{n=0}^\infty \frac{\Psi_n^*(y) \Psi_n(x)}{s + E_n}$ is the Laplace transform of $G_\tau(x, y)$ wrt τ . One thus recognizes that $\tilde{G}_s(x, y)$ is the Green's function satisfying

$$[H_\lambda + s] \tilde{G}_s(x, y) = \delta(x - y), \quad (66)$$

such as $\tilde{G}_s(x, 0) = \tilde{G}_s(0, y) = 0$. To compute $\tilde{G}_s(x, y)$, we look for two complementary functions $u_s(y)$ and $v_s(y)$ solution of the homogeneous equation $[H_\lambda + s] \psi(y) = 0$ with $u_s(0) = 0$ and $v_s(y \rightarrow \infty) = 0$. From $u_s(y)$ and $v_s(y)$, we can compute $\tilde{G}_s(x, y)$ as [42]

$$\tilde{G}_s(x, y) = \begin{cases} \frac{2}{W} u_s(x) v_s(y) & \text{if } x \leq y, \\ \frac{2}{W} u_s(y) v_s(x) & \text{if } x \geq y, \end{cases} \quad (67)$$

where

$$W = u'_s(x) v_s(x) - u_s(x) v'_s(x) \quad (68)$$

is the Wronskian associated to $u_s(x)$ and $v_s(x)$, which is here independent of x . Finally we obtain from (65) and (67)

$$\tilde{\varphi}(s) = \int_0^\infty e^{-s\tau} \varphi(\tau) d\tau = \frac{\sqrt{2}}{W} \int_0^\infty dy_F u'_s(0) v_s(y_F). \quad (69)$$

This formula (69) together with (56) and (57) allows to compute $\langle e^{-\lambda \int_0^t V(x_{\max} - x(\tau)) d\tau} \rangle$ for any function $V(x)$. To obtain explicit results from these general formulas, we need to analyze in more detail the Schrödinger operator in Eq. (59). This will be done, for some special cases, in section 4.2.

4.1.2 Brownian bridge

In the case of a BB, the method presented above can be straightforwardly adapted to compute the Laplace transform of the PDF of functional of the maximum of the Brownian bridge, $\mathcal{O}_{\max, B}(t)$, namely $\langle e^{-\lambda \int_0^t V(x_{\max, B} - x_B(\tau)) d\tau} \rangle$. In principle, one could use the same reasoning as before, i.e. break the time interval $[0, t]$ into $[0, t_{\max}]$ and $[t_{\max}, t]$ where t_{\max} is the time at which the maximum is reached (see Fig. 8). The main difference is that, for the BB, the PDF of t_{\max} is uniform $P(t_{\max}) = 1/t$, and not given by the arcsine law (53) – this is a consequence of periodic boundary conditions in the time direction. There is however a simpler way to proceed, which makes use of the Vervaat's construction (see Fig. 6), which allows us to map any functional of the maximum of a BB onto a (standard) functional of the Brownian excursion $x_E(\tau)$, from which path-integral techniques have already been developed [8, 19, 20]. Hence, generalizing the relation in (35) to more general functionals, we have

$$\int_0^t V(x_{\max, B} - x_B(\tau)) d\tau \stackrel{\text{law}}{=} \int_0^t V(x_E(\tau)) d\tau. \quad (70)$$

Hence, using that identity in law (70), the LT of the PDF of any functional of the maximum of the BB, can be written as a path-integral [8, 19, 20]. Because we are dealing with a Brownian excursion, which prevents us to impose simultaneously $x_E(0) = 0$ and $x_E(0^+) > 0$, this path integral method needs to be suitably regularized. This can be done, as explained before [see Eq. (39)], by using a

cutoff ε such that $x_E(0) = x_E(t) = \varepsilon$ and computing the statistics of these observables (70) by using a limiting procedure:

$$\langle e^{-\lambda \int_0^t V(x_{\max, B} - x_B(\tau)) d\tau} \rangle = \lim_{\varepsilon \rightarrow 0} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle}, \quad (71)$$

where H_λ is the Schrödinger operator defined above (59). Note that the denominator in (71) does not depend on the functional at hand, i.e. it is independent on $V(x)$, and it is readily computed from (61) to be

$$\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle \sim \varepsilon^2 \sqrt{\frac{2}{\pi t^3}}, \quad \varepsilon \rightarrow 0. \quad (72)$$

This formula (71) is quite general and it allows us to compute the statistics of a wide class of functionals of the maximum of the Brownian bridge, as we illustrate it below.

4.2 Applications to some specific functionals of the maximum

In this section, we present some concrete applications of the above path integral formalism, both for the free BM as well as for the BB. We first illustrate the method on the computation of the full statistics of the DOS, providing a detailed derivation of the results recently announced in Ref. [29]. Then, we study the special family of functionals corresponding to $V(x) = x^\alpha$, whose applications were discussed in the introduction. Finally, we provide a full detailed study of the special case $\alpha = -1$, i.e. $V(x) = 1/x$, which corresponds to the analysis of Odlyzko's algorithm.

4.2.1 Full statistics of the DOS

The main body of results for the statistics of the DOS [see e.g., Eqs. (82) and (83) below] were recently announced by us in a short Letter [29]. In this section we provide a detailed derivation of these results using the path-integral framework presented above.

The case of free BM. To compute the statistics of $\rho(r, t) = \int_0^t \delta(x_{\max} - x(\tau) - r) d\tau$, we apply the above formalism (56, 57, 69) to the special case $V(x) = \delta(x - r)$. In this case, the two independent solutions are $u_s(x), v_s(x)$ of $[H_\lambda + s]\psi(x) = 0$, with $H_\lambda = -(1/2)d^2/dx^2 + \lambda\delta(x - r) + V_{\text{wall}}(x)$ which reads simply here, for $x \neq r$:

$$-\frac{1}{2}\psi''(x) + s\psi(x) = 0, \text{ for } x \in [0, r[\cup]r, +\infty[\quad (73)$$

with the following boundary conditions:

$$\begin{cases} u_s(0) = 0 \\ u_s(r^+) = u_s(r^-) \\ \frac{1}{2}(u'_s(r^+) - u'_s(r^-)) = \lambda u_s(r) \end{cases}, \quad \begin{cases} \lim_{y \rightarrow \infty} v_s(y) = 0 \\ v_s(r^+) = v_s(r^-) \\ \frac{1}{2}(v'_s(r^+) - v'_s(r^-)) = \lambda v_s(r) \end{cases}, \quad (74)$$

where the boundary conditions in r results from the presence of the delta peak in $H_\lambda = -\frac{1}{2}d^2/dx^2 + \lambda\delta(x - r)$. Hence, $u_s(x)$ and $v_s(x)$ are given by

$$u_s(x) = \begin{cases} A \sinh(\sqrt{2s}x) & x \leq r \\ A \left(\sinh(\sqrt{2s}x) + \frac{2\lambda \sinh(\sqrt{2s}r)}{\sqrt{2s}} \sinh(\sqrt{2s}(x - r)) \right) & x > r \end{cases} \quad (75)$$

$$v_s(x) = \begin{cases} B \left(e^{-\sqrt{2s}(x-r)} + \frac{2\lambda}{\sqrt{2s}} \sinh(\sqrt{2s}(r - x)) \right) & x \leq r \\ B e^{-\sqrt{2s}(x-r)} & x > r, \end{cases} \quad (76)$$

where A and B are normalization constants which are irrelevant here. The Wronskian is thus given by

$$W = u'_s(x)v_s(x) - u_s(x)v'_s(x) = AB e^{\sqrt{2s}r} \left[(1 - e^{-2\sqrt{2s}r}) \lambda + \sqrt{2s} \right]. \quad (77)$$

With these formulas (65, 69, 75, 76, 77), we finally find

$$\tilde{\varphi}(s) = \frac{1}{\sqrt{s}} \frac{\sqrt{2s} + \lambda \left(1 - e^{-\sqrt{2sr}}\right)^2}{\sqrt{2s} + \lambda \left(1 - e^{-2\sqrt{2sr}}\right)}, \quad (78)$$

which, combined with Eq. (56), yields finally the formula :

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \rho(r,t)} \rangle = \frac{1}{s} \left(\frac{\sqrt{2s} + \lambda (1 - e^{-\sqrt{2sr}})^2}{\sqrt{2s} + \lambda (1 - e^{-2\sqrt{2sr}})} \right)^2. \quad (79)$$

From this expression (79), one can obtain the moments of arbitrary order $\langle \rho^k(r,t) \rangle$, for $k \in \mathbb{N}$ by expanding it in powers of λ . It yields:

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \rho(r,t)} \rangle = 1 + \sum_{k=1}^\infty (-\lambda)^k \frac{8e^{-\sqrt{2sr}}}{(2s)^{\frac{k}{2}+1}} (1 - e^{-\sqrt{2sr}})^2 (1 + ke^{-\sqrt{2sr}}) (1 - e^{-2\sqrt{2sr}})^{k-2}. \quad (80)$$

This expansion (80) in powers of λ yields the LT of the moments $\langle \rho^k(r,t) \rangle$ wrt t as:

$$\int_0^\infty \langle \rho^k(r,t) \rangle e^{-st} dt = \frac{8k!}{(2s)^{\frac{k}{2}+1}} \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} \left((2l-1)e^{-(2l+1)\sqrt{2sr}} + (k-2(l+1))e^{-2(l+1)\sqrt{2sr}} \right). \quad (81)$$

It is then possible to invert this LT using the functions $\Phi^{(j)}$'s presented in Appendix A to obtain

$$\langle \rho^k(r, t=1) \rangle = 8k! \sum_{l=0}^{k-1} (-1)^l \binom{k-1}{l} [(2l+1)\Phi^{(k+1)}((2l+1)r) + (k-2(l+1))\Phi^{(k+1)}(2(l+1)r)]. \quad (82)$$

For $k=1$, this yields back the result obtained in (32).

By inverting the LT wrt λ in (79), we obtain the full PDF $P_t(\rho, r)$ of the DOS (1), as a function of ρ , for different values of the parameter r .

$$\int_0^\infty e^{-st} P_t(\rho, r) dt = \delta(\rho) \frac{(e^{-\sqrt{2sr}} - 1)^2}{s(1 + e^{-\sqrt{2sr}})^2} + \frac{e^{-\frac{\rho\sqrt{2s}e^{\sqrt{2sr}}}{2\sinh(\sqrt{2sr})}}}{\cosh^3\left(\frac{r\sqrt{2s}}{2}\right)} \left(\frac{e^{\frac{r\sqrt{2s}}{2}}}{\sqrt{2s}} + \frac{\rho e^{\sqrt{2sr}}}{4\sinh(r\sqrt{2s})\sinh\left(\frac{r\sqrt{2s}}{2}\right)} \right). \quad (83)$$

It has an unusual form with a peak $\propto \delta(\rho)$ at $\rho=0$, in addition to a non trivial continuous background density $p_t(\rho, r)$ for $\rho > 0$. Hence one has

$$P_t(\rho, r) = F_W(r, t)\delta(\rho) + p_t(\rho, r), \quad (84)$$

where $F_W(r, t) = \text{Prob}[W(t) \leq r]$, given in (85), is the probability that the width $W(t) = \max_{\tau \in [0, t]} x(\tau) - \min_{\tau \in [0, t]} x(\tau)$ is smaller than r . This can be understood because if $W(t)$ is smaller than r , the amount of time spent by the process at a distance within $[r, r+dr]$ from the maximum is 0 (see Fig. 2), yielding the delta peak at $\rho=0$. Indeed we can check that the coefficient of the term $\propto \delta(\rho)$ (83) is the LT wrt t of

$$F_W(r, t) = 1 + \sum_{l=1}^\infty 4l(-1)^l \text{erfc}(lr/\sqrt{2t}), \quad (85)$$

which corresponds precisely to the distribution of the width (or the span) of BM [43,44].

On the other hand, in (84), $p_t(\rho, r) = p_1(\rho/\sqrt{t}, r/\sqrt{t})/\sqrt{t}$ is a regular function of ρ , for $r > 0$ (see Fig. 9), and has a more complicated structure. We obtain an explicit expression of its LT wrt t given by the second term of Eq. (83). In Fig. 9 we show the results of $p_1(\rho, r)$ obtained from numerical simulations (averages are performed over 10^7 samples) for three different values of r . We see that they are in perfect agreement with our exact formula (83). The analysis of the distribution $p_1(\rho, r)$,

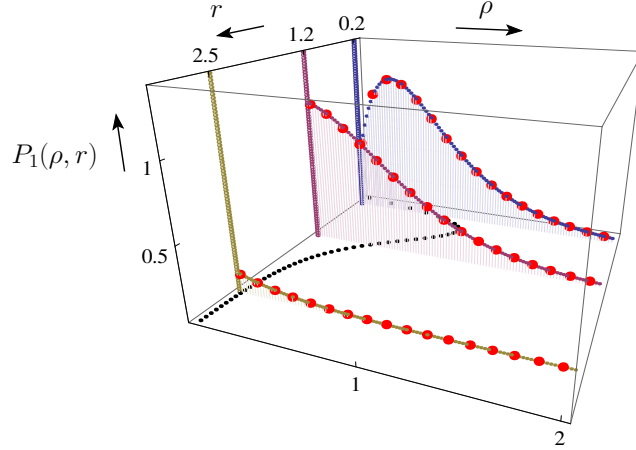


Fig. 9 (Color online) Plot of $P_1(\rho, r)$ as a function of ρ for different values of r . The solid lines for $\rho = 0$ represent the $\delta(\rho)$ peak, $\propto \delta(\rho)$ in Eq. (83). The dotted lines correspond to our exact analytical results for the background density $p_1(\rho, r)$ in (83) – where the inverse LT wrt s has been performed numerically (in green for $r = 2.5$, purple for $r = 1.2$ and blue for $r = 0.2$) – while the red dots indicate the results of simulations. On the $z = 0$ plane, we have plotted the exact mean DOS in Eq. (32).

including its asymptotic behaviors when $\rho \rightarrow 0$ and $\rho \rightarrow \infty$ was carried out in Ref. [29] and we refer the interested reader to the supplementary material of Ref. [29] for more details.

The case of a Brownian bridge. In this case, we remind that, as it can be seen using the Vervaat's construction, the DOS for the BB coincides with the DOS for the Brownian excursion, which in turn, coincides also with the local time of the excursion [see the identity in law in Eq. (35)]. The full PDF of the $\rho_B(r, t)$ can be obtained from the above general formula (71), applied to the special case $V(x) = \delta(x - r)$:

$$\langle e^{-\lambda \rho_B(r, t)} \rangle = \lim_{\varepsilon \rightarrow 0} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle}, \quad H_\lambda = -\frac{1}{2} \frac{d^2}{dx^2} + \lambda \delta(x - r) + V_{\text{wall}}(x), \quad (86)$$

where we recall that $V_{\text{wall}}(x)$ is a hard-wall potential, $V_{\text{wall}}(x) = 0$ for $x \geq 0$ and $V_{\text{wall}}(x) = +\infty$ for $x < 0$ – hence H_λ is defined on $[0, +\infty[$. The numerator in (86) can be computed by expanding the matrix element on the eigenfunctions $\psi_k(x)$ of H_λ which are given by

$$H_\lambda \psi_k(x) = \frac{k^2}{2} \psi_k(x), \quad \psi_k(x) = \begin{cases} 0 & x \leq 0 \\ a(k, \lambda) \sin(kx) & 0 \leq x \leq r \\ a(k, \lambda) \left(\sin(kx) + \frac{2\lambda}{k} \sin(kr) \sin(k(x-r)) \right) & r \leq x, \end{cases} \quad (87)$$

where the amplitude $a(k, \lambda)$ ensures the normalization condition:

$$|a(k, \lambda)|^{-2} = \frac{\pi}{2} \left[1 + \frac{4\lambda}{k} \sin(kr) \left(\cos(kr) + \frac{\lambda}{k} \sin(kr) \right) \right]. \quad (88)$$

We finally obtain

$$\langle e^{-\lambda \rho_B(r, t)} \rangle = \sqrt{\frac{2t^3}{\pi}} \int_0^\infty \frac{k^2 e^{-\frac{k^2 t}{2}}}{1 + \frac{4\lambda}{k} \sin(kr) \left(\frac{\lambda}{k} \sin(kr) + \cos(kr) \right)} dk. \quad (89)$$

Note that from the identity in law in (35) this expression (89) yields also an interesting relation for the local time of the Brownian excursion, which we have not seen in the literature. Furthermore, by expanding this formula (89) in powers of λ , it is possible to obtain, using yet another method, the moments $\langle [\rho_B(r, t)]^k \rangle$ of arbitrary order $k \in \mathbb{N}$ and recover the expression given above (42).

On the other hand, by studying the large λ behavior of (89), which is of order $\mathcal{O}(\lambda^0)$, we can show that the PDF of $\rho_B(r, t)$ has an expression similar to, albeit different from, the one for BM in (84):

$$P_{t,B}(\rho, r) = F_{W,B}(r, t)\delta(\rho) + p_{t,B}(\rho, r) , \quad (90)$$

where $F_{W,B}(r, t)$ is the distribution function of the width of the BB [45]:

$$F_{W,B}(r, t) = 1 + 2 \sum_{l=1}^{\infty} e^{-\frac{2l^2 r^2}{t}} \left(1 - \frac{4l^2 r^2}{t}\right) , \quad (91)$$

while $p_{t,B}(\rho, r)$ is now a different distribution. In particular, one can show that (42) yields back the complicated though explicit formula for $p_{t,B}(\rho, r)$ found in Ref. [46, 39] using a completely different method.

$$p_{t,B}(\rho, r) = 1 - 2 \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \binom{j-1}{k} e^{-(\rho+2rj)^2/2} (-\rho)^k H_{k+2}(\rho+2rj)/k! , \quad (92)$$

where the H_n 's are Hermite polynomials defined by

$$H_n(x) = n! \sum_{i=0}^{\lfloor n/2 \rfloor} \frac{(-1)^i x^{n-2i}}{2^i i! (n-2i)!} . \quad (93)$$

We refer the interested reader to Ref. [29] for more details on the distribution $p_{t,B}(\rho, r)$, including its asymptotic behaviors.

4.2.2 The case $V(x) = x^\alpha$: first and second moments of the functionals

We now apply this path integral formalism to the functionals $T_\alpha(t)$, for free BM (12), and $T_{\alpha,B}(t)$ the associated observable for the BB (13). We showed previously that the first moment can be obtained directly from the corresponding average DOS. This simple method can not be easily adapted to compute higher moments of $T_\alpha(t)$ or $T_{\alpha,B}(t)$. We show here how to compute these moments using the path integral formalism developed above. We will treat separately the case of free BM and BB.

The case of free BM. It is convenient to start from the following formula, obtained from the combination of Eqs. (56, 57, 65):

$$\int_0^\infty dt e^{-st} \langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \frac{1}{2} \left(\int_0^\infty dy \partial_x \tilde{G}_s(0, y) \right)^2 , \quad (94)$$

where we recall that $\tilde{G}_s(x, y)$ is the Green's function of the following Schrödinger equation

$$\left[-\frac{d^2}{dx^2} + \lambda V(x) + V_{\text{wall}}(x) + s \right] \tilde{G}_s(x, y) = \delta(x - y) . \quad (95)$$

Of course, there exist very few instances of potential $V(x)$ for which this Eq. (95) can be solved exactly (see below). However, the moments of $T_\alpha(t)$ can be extracted, in principle, for a generic $V(x)$. Indeed, from (94), we see that the moments of $T_\alpha(t)$ are obtained by successive derivations of the right hand side of (94) wrt λ , evaluated in $\lambda = 0$. These successive derivatives can in turn be expressed as combinations of the successive derivatives of $\tilde{G}_s(x, y)$ wrt λ . Indeed, if we write the following expansion:

$$\tilde{G}_s(x, y) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} f_n(x, y) , \quad (96)$$

one has from (94):

$$\int_0^\infty \langle T_\alpha^k(t) \rangle e^{-st} dt = (-1)^k \frac{k!}{2} \sum_{n=0}^k b_n b_{k-n} , \quad b_n = \frac{1}{n!} \int_0^\infty \partial_x f_n(x, y)|_{x=0} dy . \quad (97)$$

Hence, we need to solve perturbatively the equation for the Green's function (95) in powers of λ to compute the functions $f_n(x, y)$ in (96). First it is easy to obtain $f_0(x, y)$ as

$$f_0(x, y) = \frac{1}{\sqrt{2s}} (e^{-\sqrt{2s}|x-y|} - e^{-\sqrt{2s}(x+y)}). \quad (98)$$

Furthermore, one can show that the functions $f_n(x, y)$ satisfy the following recursion relation:

$$f_n(x, y) = -n \int_0^\infty dz f_0(x, z) V(z) f_{n-1}(x, y), \quad n \geq 1. \quad (99)$$

The recursion relation (99) can be solved formally in the closed form

$$f_n(x, y) = (-1)^n n! \prod_{i=1}^n \int_0^\infty dz_i f_0(x, z_i) \prod_{i=1}^{n-1} V(z_i) f_0(z_i, z_{i+1}) V(z_n) f_0(z_n, y). \quad (100)$$

Hence, from Eq. (97) together with (100) one can compute the moments of arbitrary order of any functional of the maximum of the free BM. Note that this technique is a generalization of the method based on propagators which we used before for the computation of the average DOS, corresponding to a particular potential $V(x) = \delta(x - r)$. Here we apply this formalism to the case $V(x) = x^\alpha$.

We first compute the first moment, using Eq. (97) for $k = 1$, together with (100). We obtain:

$$\int_0^\infty dt e^{-st} \langle T_\alpha(t) \rangle = - \left(\int_0^\infty dy \partial_x f_1(x, y) \Big|_{x=0} \right) \left(\int_0^\infty dy \partial_x f_0(x, y) \Big|_{x=0} \right). \quad (101)$$

Using the expressions of $f_0(x, y)$ in (98) and of $f_1(x, y)$, obtained from Eq. (100) together with the fact that $\partial_x f_0(0, y) = 2e^{-\sqrt{2s}y}$, we obtain (performing the change of variable $z \rightarrow \sqrt{2s}z$):

$$\begin{aligned} \int_0^\infty dy \partial_x f_1(0, y) &= - \int_0^\infty dy \int_0^\infty dz \partial_x f_0(0, z) z^\alpha f_0(z, y) \\ &= - \frac{2}{\sqrt{2s}^{\alpha+3}} \int_0^\infty dy \int_0^\infty dz e^{-z} z^\alpha (e^{-|z-y|} - e^{-(z+y)}) \\ &= - \frac{2}{(2s)^{\frac{\alpha+3}{2}}} \Gamma(\alpha+1) (2 - 2^{-\alpha}). \end{aligned} \quad (102)$$

Hence, combining (101, 102) and using that $\int_0^\infty dy \partial_x f_0(x, y) \Big|_{x=0} = \sqrt{2/s}$ we arrive at

$$\int_0^\infty dt e^{-st} \langle T_\alpha(t) \rangle = \frac{4}{(2s)^{\frac{\alpha+4}{2}}} \Gamma(\alpha+1) (2 - 2^{-\alpha}), \quad (103)$$

which after Laplace inversion, yields

$$\langle T_\alpha(t) \rangle = \frac{(2t)^{1+\frac{\alpha}{2}} (2 - 2^{-\alpha}) \Gamma(\frac{1+\alpha}{2})}{(2+\alpha)\sqrt{\pi}}. \quad (104)$$

This formula coincides, as it should, with the formula obtained directly from the average DOS in (33).

We now compute the second moment, using Eq. (97) for $k = 2$:

$$\int_0^\infty dt e^{-st} \langle T_\alpha^2(t) \rangle = \left(\int_0^\infty dy \partial_x f_1(0, y) \right)^2 + \left(\int_0^\infty dy \partial_x f_0(0, y) \right) \left(\int_0^\infty dy \partial_x f_2(0, y) \right). \quad (105)$$

We then have, using (98, 100), and performing the change of variables with the substitution $u \rightarrow \sqrt{2s} z_1$, $v \rightarrow \sqrt{2s} z_2$:

$$\begin{aligned} \int_0^\infty dy \partial_x f_2(0, y) &= 2 \int_0^\infty dy \int_0^\infty dz_1 \int_0^\infty dz_2 \partial_x f_0(0, z_1) z_1^\alpha f_0(z_1, z_2) z_2^\alpha f_0(z_2, y) \\ &= \frac{4}{\sqrt{2s}^{5+2\alpha}} \int_0^\infty dy \int_0^\infty du \int_0^\infty dv e^{-u} u^\alpha (e^{-|u-v|} - e^{-(u+v)}) v^\alpha (e^{-|v-y|} - e^{-(v+y)}) \\ &= \frac{8}{\sqrt{2s}^{5+2\alpha}} C_\alpha, \quad \text{with } C_\alpha = \int_0^\infty du \int_0^\infty dv e^{-u} u^\alpha (e^{-|u-v|} - e^{-(u+v)}) v^\alpha (1 - e^{-v}) \end{aligned} \quad (106)$$

where C_α can be explicitly computed as

$$C_\alpha = \sum_{n=1}^{\infty} \frac{\Gamma(2+2\alpha+n)}{n!2^{2+2\alpha+n}(1+\alpha+n)} + \frac{(4^{1+\alpha}-1)\Gamma(3+2\alpha)}{2^{3+2\alpha}(1+\alpha)^2} - \frac{2^{1+\alpha}-1}{2^{2+2\alpha}}\Gamma(1+\alpha)^2. \quad (107)$$

Finally, using (98, 102, 105, 106), we obtain

$$\int_0^\infty dt e^{-st} \langle T_\alpha^2(t) \rangle = \frac{4}{(2s)^{3+\alpha}} (\Gamma(\alpha+1)^2(2-2^{-\alpha})^2 + 4C_\alpha), \quad (108)$$

which after Laplace inversion yields:

$$\begin{aligned} \langle T_\alpha^2(t) \rangle = \frac{t^{2+\alpha}}{2^{3\alpha}\Gamma(3+\alpha)} & \left(\Gamma(\alpha+1)^2(2^\alpha-1)(2^{\alpha+1}-1) + \frac{\Gamma(3+2\alpha)(4^{\alpha+1}-1)}{4(1+\alpha)^2} \right. \\ & \left. + \sum_{n=1}^{\infty} \frac{\Gamma(2+2\alpha+n)}{n!2^{1+n}(1+\alpha+n)} \right). \end{aligned} \quad (109)$$

The last sum over n can be finally expressed in terms of incomplete beta function, which yields the formula (16). In particular, by taking (carefully) the limit $\alpha \rightarrow -1$ in the above expression (109), we recover the result of [33]

$$\lim_{\alpha \rightarrow -1} \langle T_\alpha^2(t=1) \rangle = \frac{\pi^2}{3} + 4\log(2)^2, \quad (110)$$

which was obtained by the authors of [33] using a completely different method.

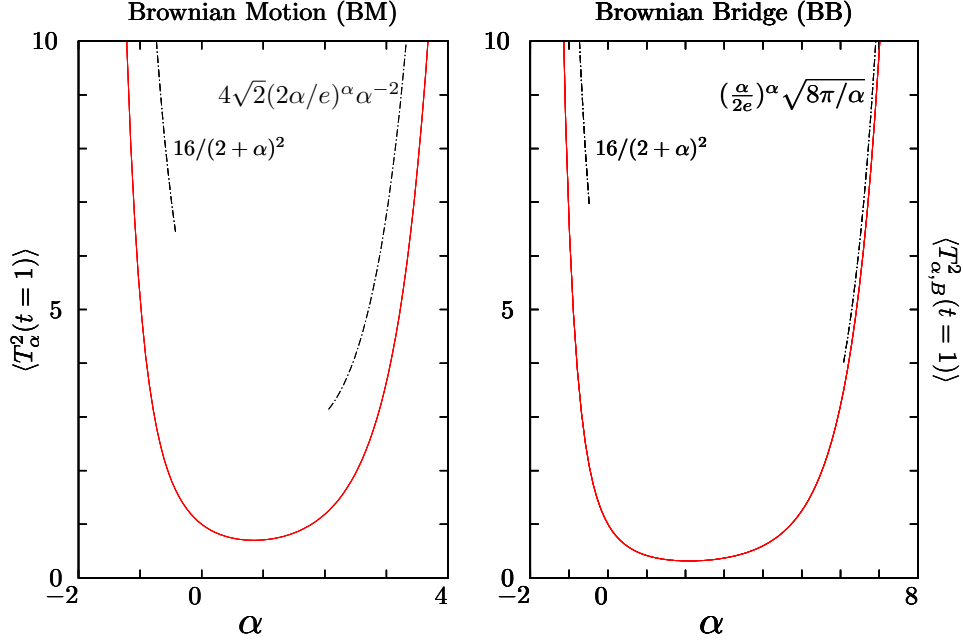


Fig. 10 **Left panel:** Plot of $\langle T_\alpha^2(t=1) \rangle$, as a function of α , for the BM, as given in Eq. (109). **Right panel:** Plot of $\langle T_{\alpha,B}^2(t=1) \rangle$, as a function of α , for the BB, as given in Eq. (121). In both panels, the dashed black lines indicate the asymptotic behaviors discussed in the introduction. In particular, for $\alpha = -1$ for the BM (corresponding to Odlyzko's algorithm), we recover $\langle T_{\alpha=-1}^2(t=1) \rangle = \frac{\pi^2}{3} + 4\log(2)^2$ [32]. For $\alpha = 1$ for the BB (corresponding to the Airy distribution), we recover $\langle T_{\alpha=1,B}^2(t=1) \rangle = \frac{5}{12}$ [19, 20].

The case of BB. In this case, the equivalent of Eq. (94) is the formula derived in Eqs. (71) and (72) which reads

$$\langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \sqrt{\frac{\pi}{2}} t^{3/2} \partial_x \partial_y G_t(0, 0). \quad (111)$$

Taking the LT of the above relation wrt t , one obtains:

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle e^{-\lambda \int_0^t d\tau V(x_{\max} - x(\tau))} \rangle = \sqrt{\frac{\pi}{2}} \partial_x \partial_y \tilde{G}_s(x, y) \Big|_{x=y=0}, \quad (112)$$

where $\tilde{G}_s(x, y)$ is the Green's function of the operator in (95). From (112), one deduces the following relation

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle \int_0^t V(x_{\max, B} - x_B(\tau)) d\tau \rangle^k = (-1)^k \sqrt{\frac{\pi}{2}} \frac{\partial^2}{\partial x \partial y} f_k(x, y) \Big|_{x=y=0}, \quad (113)$$

where $f_k(x, y)$ is given in Eq. (100). From (113), specified to the case $V(x) = x^\alpha$ and $k = 1$, we obtain the first moment $\langle T_{\alpha, B}(t) \rangle$ as

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha, B}(t) \rangle = -\sqrt{\frac{\pi}{2}} \frac{\partial^2}{\partial x \partial y} f_1(x, y) \Big|_{x=y=0}. \quad (114)$$

Hence, using (98, 100) together with the change of variable $z \rightarrow \sqrt{2s}z$, we obtain

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha, B}(t) \rangle = \sqrt{\frac{\pi}{2}} \int_0^\infty dz \partial_x f_0(0, z) z^\alpha \partial_y f_0(z, 0) \quad (115)$$

$$= \sqrt{\frac{\pi}{2}} \frac{4}{\sqrt{2s}^{\alpha+1}} \int_0^\infty dz e^{-2z} z^\alpha = \sqrt{\frac{\pi}{2}} \frac{2^{1-\alpha}}{\sqrt{2s}^{\alpha+1}} \Gamma(1 + \alpha), \quad (116)$$

which, after Laplace inversion, yields immediately

$$\langle T_{\alpha, B}(t) \rangle = \frac{t^{1+\alpha/2}}{2^{\alpha/2}} \Gamma\left(1 + \frac{\alpha}{2}\right), \quad (117)$$

which coincides with the formula obtained above from the average DOS (44).

We can also compute the second moment $\langle T_{\alpha, B}^2(t) \rangle$ by using Eq. (112) for the case $k = 2$, yielding

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha, B}^2(t) \rangle = \sqrt{\frac{\pi}{2}} \partial_x \partial_y f_2(0, 0), \quad (118)$$

where $f_2(x, y)$ can be obtained from (100) with $V(x) = x^\alpha$. Using (98, 100), we obtain, performing the changes of variables $u \rightarrow \sqrt{2s}z_1, v \rightarrow \sqrt{2s}z_2$, and after some manipulations

$$\int_0^\infty dt e^{-st} t^{-3/2} \langle T_{\alpha, B}^2(t) \rangle = \sqrt{2\pi} \int_0^\infty dz_1 \int_0^\infty dz_2 \partial_x f_0(0, z_1) z_1^\alpha f_0(z_1, z_2) z_2^\alpha \partial_y f_0(z_2, 0) \quad (119)$$

$$= \frac{\sqrt{2\pi}}{2^{2\alpha} \sqrt{2s}^{2\alpha+3}} \left(\frac{2\Gamma(2+2\alpha)}{1+\alpha} - \Gamma(1+\alpha)^2 \right), \quad (120)$$

which, after inverse Laplace transform, yields

$$\langle T_{\alpha, B}^2(t) \rangle = t^{\alpha+2} \frac{\sqrt{\pi}}{2^{3\alpha+1} \Gamma(\alpha + \frac{3}{2})} \left(\frac{2\Gamma(2+2\alpha)}{1+\alpha} - \Gamma(1+\alpha)^2 \right). \quad (121)$$

In particular, we can check that $\langle T_{\alpha=1, B}^2(t=1) \rangle = \frac{5}{12}$ coincide with the known result for the area under a Brownian excursion, i.e. the second moment of the Airy distribution [19, 20].

4.2.3 The exactly solvable case $V(x) \propto 1/x$ and application to the Odlyzko's algorithm

In this section, we apply our general formalism to the special case $V(x) = 1/x$, which is relevant to analyze the distribution of the cost of the optimal algorithm to find the maximum of a random walk (see Eq. (4)), as shown in Refs. [30, 33]. We apply this formalism separately both to the free Brownian motion and to the Brownian bridge.

The case of free BM. In this case, we want to construct the Green's function $\tilde{G}_s(x, y)$ in (67) which can be built from two independent solutions $u_s(x)$ and $v_s(x)$ of the following Schrödinger equation

$$\left[-\frac{1}{2} \frac{d^2}{dx^2} + \frac{\lambda}{x} + s \right] \psi(x) = 0, \quad s > 0 \quad (122)$$

with the asymptotic behaviors

$$u_s(0) = 0 \quad \& \quad v_s(y \rightarrow \infty) = 0. \quad (123)$$

It turns out that the above Schrödinger equation (122) can be solved in terms of hypergeometric functions. The solutions $u_s(x)$ and $v_s(x)$ satisfying (123) read

$$u_s(x) = \mathcal{A} e^{-\sqrt{2s}x} x {}_1F_1 \left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}x \right), \quad (124)$$

$$v_s(x) = \mathcal{B} e^{-\sqrt{2s}x} x U \left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}x \right), \quad (125)$$

where ${}_1F_1(a, b, x)$ and $U(a, b, x)$ are confluent hypergeometric functions (respectively Kummer's and Tricomi's hypergeometric function) and where \mathcal{A} and \mathcal{B} are two irrelevant numerical constants (note that $u_s(x)$ and $v_s(x)$ are also known as Whittaker functions). The Wronskian W reads

$$W = u'_s(x)v_s(x) - u_s(x)v'_s(x) = \frac{\mathcal{A}\mathcal{B}}{2\lambda\Gamma(\lambda/\sqrt{2s})}. \quad (126)$$

Note that one can check from (124) that $u_s(0) = 0$, as it should (123) and that $u'_s(0) = \mathcal{A}$. The function $\tilde{\varphi}(s)$ in Eq. (69) reads in this special case (124, 125):

$$\tilde{\varphi}(s) = 2^{3/2}\lambda\Gamma\left(\frac{\lambda}{\sqrt{2s}}\right) \int_0^\infty e^{-\sqrt{2s}y} y U\left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}y\right) dy. \quad (127)$$

After some manipulations, this integral (127) can be evaluated as (see Appendix D.1)

$$\begin{aligned} \tilde{\varphi}(s) &= \frac{1}{\sqrt{s}} G\left(\frac{\lambda}{\sqrt{2s}}\right), \\ G(x) &= 2 \sum_{k=0}^{\infty} (-1)^{k+1} \tilde{\zeta}(k) x^k, \quad \text{with } \tilde{\zeta}(k) = (1 - 2^{-1-k}) \zeta(k) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^k}. \end{aligned} \quad (128)$$

Hence, from the general formulae (56, 57), together with (128), one obtains the explicit formula for the case $V(x) = 1/x$:

$$\int_0^\infty e^{-st} \langle e^{-\lambda \int_0^t \frac{d\tau}{x_{\max-x(\tau)}}} \rangle dt = \frac{4}{s} \sum_{n=0}^{\infty} (-1)^n \frac{\lambda^n}{(\sqrt{2s})^n} \sum_{k=0}^n \tilde{\zeta}(k) \tilde{\zeta}(n-k). \quad (129)$$

Expanding the left hand side of Eq. (129) in powers of λ we obtain the moments of $T_{\alpha=-1}(t=1)$ of arbitrary order, as announced in the introduction (18),

$$\langle T_{\alpha=-1}^k(t=1) \rangle = \Gamma\left(\frac{k+1}{2}\right) \frac{2^{\frac{k}{2}+2}}{\sqrt{\pi}} \sum_{m=0}^k \tilde{\zeta}(m) \tilde{\zeta}(k-m). \quad (130)$$

This allows us to recover in a completely different manner the results obtained in [33] by probabilistic tools. For completeness, we mention that the authors of [33] obtain an explicit expression of the PDF $p(s)$ of $T_{\alpha=-1}(t=1)$ as (see Theorem 4.2 of that paper)

$$p(s) = \frac{8}{s^3} \Theta\left(\frac{4}{s^2}\right), \quad \Theta(x) = \int_0^x \theta_1(y) \theta_2(x-y) dy, \quad (131)$$

where the functions $\theta_1(x)$ and $\theta_2(x)$ are given by

$$\theta_1(x) = \sum_{n=1}^{\infty} \int_x^{\infty} \frac{du}{u} \exp\left(-\tilde{n}^2 \frac{u}{2}\right), \quad \tilde{n} = \pi\left(n - \frac{1}{2}\right), \quad (132)$$

$$\theta_2(x) = \frac{\partial}{\partial x} \sum_{n=-\infty}^{+\infty} (1 - n^2 \pi^2 x) e^{-n^2 \pi^2 x/2}. \quad (133)$$

The case of the BB. In this case, the starting point of our analysis is the general formula given in (71):

$$\left\langle \exp\left[-\lambda \int_0^t \frac{d\tau}{(x_{\max,B} - x_B(\tau))}\right] \right\rangle = \lim_{\varepsilon \rightarrow 0} \frac{\langle \varepsilon | e^{-H_\lambda t} | \varepsilon \rangle}{\langle \varepsilon | e^{-H_0 t} | \varepsilon \rangle}, \quad (134)$$

where H_λ is the Schrödinger operator defined in (59), which reads here

$$H_\lambda = -\frac{1}{2} \frac{d^2}{dx^2} + \frac{\lambda}{x} + V_{\text{wall}}(x), \quad \lambda \geq 0, \quad (135)$$

while the small ε behavior of the denominator of (134) is given in (72). To compute the numerator in Eq. (134), we expand the matrix element on the eigenfunctions $|\phi_E\rangle$ of H_λ , which satisfy

$$H_\lambda |\phi_E\rangle = E |\phi_E\rangle, \quad (136)$$

where the eigenvalues $E > 0$ form a continuous spectrum, as there are no bound states here [we recall that $\lambda \geq 0$ in (135)] where the eigenvectors satisfy the boundary condition

$$\lim_{x \rightarrow 0^+} \phi_E(x) = \lim_{x \rightarrow 0^+} \langle x | \phi_E \rangle = 0. \quad (137)$$

The general solution of the Schrödinger equation (135, 136) reads

$$\phi_E(x) = c_E x e^{-i\sqrt{2E}x} {}_1F_1\left(1 - i\frac{s}{\sqrt{2E}}, 2, 2ix\sqrt{2E}\right) + d_E x e^{-i\sqrt{2E}x} U\left(1 - i\frac{s}{\sqrt{2E}}, 2, 2ix\sqrt{2E}\right), \quad (138)$$

where we recall that ${}_1F_1(a, b, x)$ and $U(a, b, x)$ are confluent hypergeometric functions. The boundary condition at $x = 0$ (137) imposes that $d_E = 0$. The remaining task is to compute the normalization constant c_E such that

$$\int_0^\infty \phi_{E'}^*(x) \phi_E(x) dx = \delta(E - E'). \quad (139)$$

Using a formula given in Landau-Lifshitz (see Appendix D.2), one can show that

$$|c_E|^2 = \frac{4\lambda}{\exp\left(\frac{2\pi\lambda}{\sqrt{2E}}\right) - 1}. \quad (140)$$

Therefore, Eq. (134), together with Eq. (140) yield

$$\left\langle \exp\left[-\lambda \int_0^t \frac{d\tau}{(x_{\max,B} - x_B(\tau))}\right] \right\rangle = 4\sqrt{\frac{\pi}{2}} t^{3/2} \lambda \int_0^\infty \frac{e^{-Et}}{(\exp[2\pi\lambda/(\sqrt{2E})] - 1)} dE. \quad (141)$$

It is then possible, using residue theorem, to invert the LT wrt λ in (141) and then perform the remaining integral over E . This finally yields the PDF of $T_{\alpha=-1,B}(t=1)$ given in Eq. (20). The expression for the moments (19) then follow straightforwardly.

5 Conclusion

In this paper, we have presented several analytical tools to study functionals of the Brownian motion and its variants, in particular Brownian bridge and Brownian excursion. These tools include (i) a “paths counting” method, relying on propagators of BM with appropriate boundary conditions and (ii) a suitably adapted path-integral method, which allows us to recast the study of functionals of the maximum of BM into the study of quantum mechanical problems. The first method (i) is conceptually quite simple and allows us to obtain in a rather simple manner the mean value of any functional of the BM, while the second method (ii) is better adapted to compute the full PDF of such functionals. We have used these methods to calculate the statistics of the density of near-extremes, or density of states (DOS), for Brownian motion $\rho(r, t)$ and its variants. In particular, from the mean DOS $\langle \rho(r, t) \rangle$, one can compute the average value of any functional of the maximum of BM. Then, we provided a thorough study of functionals of the form $T_\alpha(t) = \int_0^t (x_{\max} - x(\tau))^\alpha d\tau$, with $\alpha \in]-2, +\infty[$. As α is varied, $T_\alpha(t)$ interpolates between various physical observables, as discussed above. We have obtained an exact expression for the two first moments $\langle T_\alpha(t) \rangle$ and $\langle T_\alpha^2(t) \rangle$ both of which exhibit a non-trivial, non-monotonic behavior as a function of α . Thanks to the path-integral method, when the associated quantum problem can be solved exactly, it is possible to obtain an explicit expression of the Laplace transform for the full PDF of $T_\alpha(t)$ or $T_{\alpha,B}(t)$, from which moments of arbitrary order and in some cases the full PDF can be computed. We have worked out in detail the case $\alpha = -1$, corresponding to $V(x) \propto 1/x$, which corresponds to the cost of the optimal algorithm (due to Odlyzko’s) to find the maximum of a discrete RW. In this case, we provided an explicit expression for the moments of arbitrary order $\langle T_{\alpha=-1}^k(t) \rangle$, recovering by physical methods the results obtained in Ref. [33] by completely different probabilistic approaches. Furthermore, we have generalized these results to functionals of the Brownian bridge, $T_{\alpha,B}(t)$. In particular, we argued that, for $\alpha = -1$, the random variable $T_{\alpha=-1,B}(t)$ describes the cost of the optimal algorithm (i.e. Odlyzko’s algorithm) for the search of the maximum of a RW in a bridge configuration and computed explicitly its PDF as well as its moments of arbitrary order.

Several interesting questions are left open. For instance, here we have studied the case of a single Brownian motion and it would be interesting to extend this study to the case of multi-particle systems, where there are $N > 1$ walkers, which could be independent or instead interacting, as in the case of non-intersecting (vicious) walkers, whose extreme value statistics have recently attracted some attention [47]. Finally, we have treated the case of Brownian motion, which is continuous both in space and in time, it would be interesting to extend these results to random walks, which are discrete in time. This would, in particular, allow us to study the DOS of Lévy flights, whose behavior is qualitatively different from Brownian motion.

Acknowledgements We acknowledge support by the Indo-French Centre for the Promotion of Advanced Research under Project 4604 – 3. We acknowledge a useful correspondence with Philippe Chassaing.

A Some useful functions

We introduce the family of functions $\Phi^{(j)}$, $j \in \mathbb{N}$, which satisfy

$$\frac{e^{-\sqrt{2s}u}}{(\sqrt{2s})^{j+1}} = \int_0^\infty t^{\frac{j-1}{2}} \Phi^{(j)}\left(\frac{u}{\sqrt{t}}\right) e^{-st} dt. \quad (142)$$

These functions can be obtained explicitly by induction, using [48]

$$\Phi^{(0)}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \Phi^{(j+1)}(x) = \int_x^\infty \Phi^{(j)}(u) du. \quad (143)$$

The first functions can easily be computed as

$$\Phi^{(0)}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \quad (144)$$

$$\Phi^{(1)}(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right), \quad (145)$$

$$\Phi^{(2)}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} - \frac{1}{2} x \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right). \quad (146)$$

More generally, one can show [48] that they can be written in the form

$$\Phi^{(j)}(x) = p_j(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + q_j(x) \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right), \quad (147)$$

where $p_j(x)$ and $q_j(x)$ are rational polynomials of degree $j-2$ and $j-1$, respectively, for $j \geq 2$ [48]. We refer the interested reader to Ref. [48] for efficient algorithms, which can be implemented numerically, to compute these polynomials.

B Average DOS for reflected Brownian motion

Using the method, based on propagators, explained in section 3.1, see Eq. (27), we can also compute the average DOS for the reflected Brownian motion $x_R(\tau)$ which is the absolute value of the Brownian motion, $x_R(\tau) = |x(\tau)|$. The expression in (27), see also Fig. 4, indicates that we need to compute the propagator of the reflected Brownian motion such that $x_R(\tau) \leq M$ or equivalently $-M \leq x(\tau) \leq M$. Therefore, we compute the propagator of a Brownian particle confined in a given interval $[-M, M]$ with absorbing boundary conditions both in $x = -M$ and $x = M$. Denoting by $G_M^R(\alpha|\beta, t)$ the propagator of such a particle starting at α and ending, at time t , at β , its LT wrt t is given by

$$\tilde{G}_M^R(\alpha|\beta, s) = \frac{2 \sinh\left(\sqrt{2s}(M - \max(\alpha, \beta))\right) \sinh\left(\sqrt{2s}(M + \min(\alpha, \beta))\right)}{\sqrt{2s} \sinh\left(\sqrt{2s}2M\right)} \quad (148)$$

In order to compute the average DOS $\langle \rho^R(r, t) \rangle$ for the reflected BM, we evaluate the “number” of Brownian trajectories satisfying the following constraints: the process reaches its maximum M or their minimum $-M$ at time t_{ext} , passes through $M-r$ or $-M+r$ at time τ and end in $x_F \in [-M, M]$ at time t . The total number of such trajectories is then obtained by integrating over x_F , M and t_{ext} . When dividing the time interval $[0, t]$ into three parts delimited by τ , t_{ext} , 8 different cases may arise: $\tau < t_{\text{ext}}$ or $\tau > t_{\text{ext}}$, $X(t_{\text{ext}}) = \pm M$ and $X(\tau) = \pm(M-r)$. Using the invariance of the process under the reflection symmetry $X \rightarrow -X$ we have to consider only four different cases (each one with a multiplicity of 2):

$$\begin{aligned} \langle \rho^R(r, t) \rangle = & \lim_{\varepsilon \rightarrow 0} \frac{2}{Z_R(\varepsilon)} \int_r^\infty dM \int_0^t dt_{\text{ext}} \int_{-M}^M dx_F \\ & \left[\int_0^{t_{\text{ext}}} d\tau G_M^R(0|M-r, \tau) G_M^R(M-r|M-\varepsilon, t_{\text{ext}}-\tau) G_M^R(M-\varepsilon|x_F, t-t_{\text{ext}}) \right. \\ & + \int_{t_{\text{ext}}}^t d\tau G_M^R(0|M-\varepsilon, t_{\text{ext}}) G_M^R(M-\varepsilon|M-r, \tau-t_{\text{ext}}) G_M^R(M-r|x_F, t-\tau) \\ & + \int_0^{t_{\text{ext}}} d\tau G_M^R(0|r-M, \tau) G_M^R(r-M|M-\varepsilon, t_{\text{ext}}-\tau) G_M^R(M-\varepsilon|x_F, t-t_{\text{ext}}) \\ & \left. + \int_{t_{\text{ext}}}^t d\tau G_M^R(0|M-\varepsilon, t_{\text{ext}}) G_M^R(M-\varepsilon|r-M, \tau-t_{\text{ext}}) G_M^R(r-M|x_F, t-\tau) \right], \quad (149) \end{aligned}$$

where we have used the Markov property of BM and where $Z_R(\epsilon)$ is the normalization constant

$$Z_R(\epsilon) = 2 \int_0^\infty dM \int_0^t dt_{\text{ext}} \int_{-M}^M dx_F G_M^R(0|M-\epsilon, t_{\text{ext}}) G_M^R(M-\epsilon|x_F, t-t_{\text{ext}}). \quad (150)$$

The normalization is easily computed as $Z_R(\epsilon) \sim 2\epsilon^2$, when $\epsilon \rightarrow 0$. Using the same kind of calculations as in section 3.1 – exploiting the convolution structure of the above integrals – we find, after some manipulations,

$$\langle \rho_R(r, t=1) \rangle = 8 \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{-3-2n+12n^2+8n^3} \left(\sum_{k=0}^1 (3+(-1)^k(2n+k)^{k+1})(2n+k)^2 \Phi^{(2)}((2n+k)r) \right), \quad (151)$$

where $\Phi^{(2)}(x)$ is given in Eq. (146).

Similarly, we can study the DOS of the reflected Brownian bridge $x_{RB}(\tau)$, which is the absolute value of a Brownian bridge $x_{BB}(\tau) = |x_{BB}(\tau)|$. The calculation of the DOS in this case is very similar to the case of the free reflected BM in (149) without the integral over x_F which is set to $x_F = 0$. Using the new time reversal symmetry, we can show that the average DOS $\langle \rho_B^R(r, t) \rangle$ is given by

$$\begin{aligned} \langle \rho_{RB}(r, t) \rangle &= \lim_{\epsilon \rightarrow 0} \frac{4}{Z} \int_r^\infty dM \int_0^t dt_{\text{ext}} \int_{t_{\text{ext}}}^t d\tau \\ &\quad \left[G_M^R(0|M-\epsilon, t_{\text{ext}}) G_M^R(M-\epsilon|r-M, \tau-t_{\text{ext}}) G_M^R(r-M|0, t-\tau) \right. \\ &\quad \left. + G_M^R(0|M-\epsilon, t_{\text{ext}}) G_M^R(M-\epsilon|M-r, \tau-t_{\text{ext}}) G_M^R(M-r|0, t-\tau) \right] \end{aligned} \quad (152)$$

where $Z_{RB}(\epsilon)$ is the normalization constant, given by

$$Z_{RB}(\epsilon) = 2 \int_0^\infty dM \int_0^t dt_{\text{ext}} G_M^R(0|M-\epsilon, t_{\text{ext}}) G_M^R(M-\epsilon|0, t-t_{\text{ext}}) \quad (153)$$

The normalization is easily computed as $Z_{RB}(\epsilon) \sim 2\epsilon^2/(\sqrt{2\pi t})$, as $\epsilon \rightarrow 0$ and eventually the average DOS $\langle \rho_B^R(r, t) \rangle$ is obtained as:

$$\langle \rho_{RB}(r, t=1) \rangle = 2\sqrt{2\pi} \left(4 \sum_{n=0}^{\infty} n(-1)^{n+1} \Phi^{(1)}(2nr) - \Phi^{(1)}(2r) \right) \quad (154)$$

where $\Phi^{(1)}(x)$ is given in Eq. (145).

C Odlyzko's algorithm

C.1 Main ideas behind Odlyzko's algorithm

In this section, we give the main ideas of Odlyzko's algorithm. To get familiar with this algorithm, it is useful to consider a simpler search algorithm, denoted by u , belonging to A_n (that denotes the ensemble of the algorithms that find the maximum M_n of a random walk of n steps), which proceeds as follows: u probes always the random walk at the step where the upper envelope of the (still) possible trajectories reaches its maximum. This algorithm u is based on the idea that, as illustrated in Fig. 11, if X_m and X_{m+k} have been probed, then the searcher knows for sure that, between step m and step $m+k$, the position of the random walker can not exceed $(X_m + X_{m+k} + k)/2$. This can be shown as follows. Let us denote by n_+ the number of up-steps (+1) and n_- the number of down-steps (-1) between step m and step $m+k$. Then n_+ and n_- satisfy the equations

$$n_+ + n_- = k \quad (155)$$

$$n_+ - n_- = X_{m+k} - X_m. \quad (156)$$

Hence one has

$$n_+ = \frac{X_{m+k} - X_m + k}{2} \quad (157)$$

$$n_- = \frac{X_m - X_{m+k} + k}{2}. \quad (158)$$

Therefore the position of the random walker can not exceed $X_m + n_+ = (X_m + X_{m+k} + k)/2$, as shown in Fig. 11. This simple algorithm is illustrated in Fig. 12 on a realization of the RW for $n = 14$ steps. This basic idea

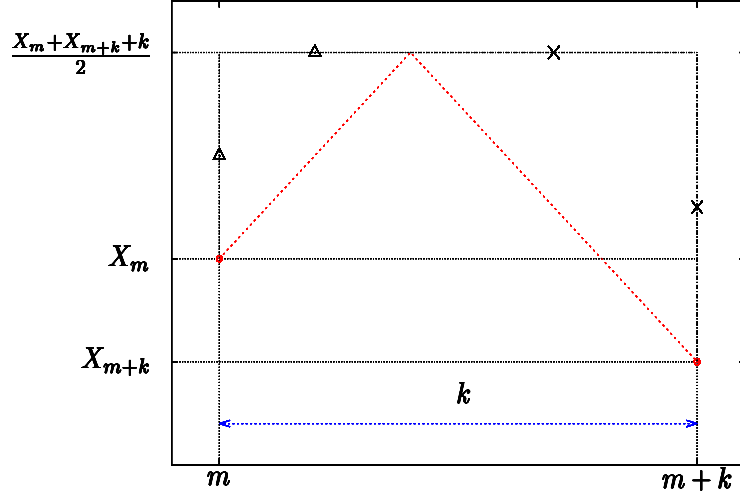


Fig. 11 Illustration of the main idea of Odlyzko's optimal algorithm. The RW can not exceed $(X_m + X_{m+k} + k)/2$ between m and $m+k$. If this quantity is smaller than $M^\#$, a new probe between m and $m+k$ is useless.

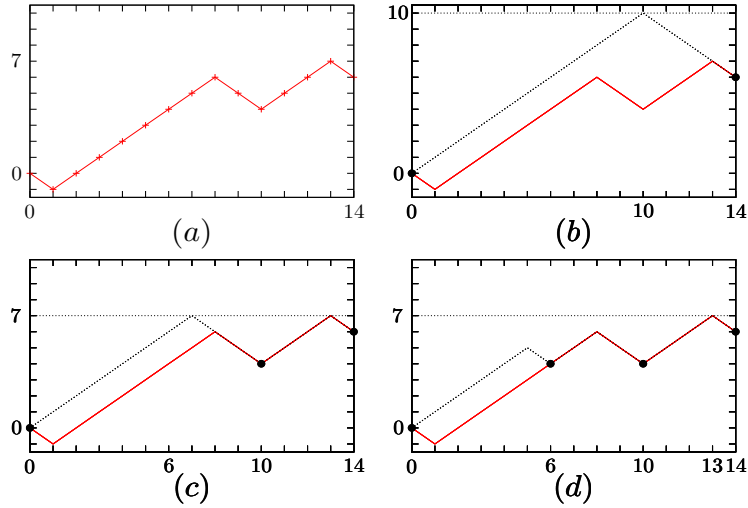


Fig. 12 An example of the algorithm u for finding $M_{14} = 7$ for a RW in 4 probes. (a) Typical realization of a 14 steps RW, for which we want to find the maximum. We know without any probe that $0 \leq M_{14} \leq 14$, and if $M_{14} = 14$, the maximum would be at position 14 (RW with only +1 jump) so we probe the position 14. (b) The first probe shows $X_{14} = 6$ and we know now (see Fig.11) that $6 \leq M_{14} \leq 10$, and if $M_{14} = 10$, the maximum would be at position 10 (dashed line) so we probe the position 10. (c) The second probe shows $X_{10} = 4$ and we know now that $6 \leq M_{14} \leq 7$, and if $M_{14} = 7$, the maximum would be at position 6 or 13 (dashed line) so we probe the position 6. (d) The third probe shows $X_6 = 4$ and we know now that $6 \leq M_{14} \leq 7$, and if $M_{14} = 7$, the maximum would be at position 13 (dashed line) so we probe position 13 and find the maximum $M_{14} = X_{13} = 7$ in 4 probes.

is at the heart of the algorithm proposed by Odlyzko.

Here we also want to explain briefly the occurrence of this particular functional of the maximum, I in (4), in the analysis of this optimal algorithm, following the line of reasoning of [32,33]. To understand this, let us consider a traveler, moving on a line, its position being denoted by y . Suppose that its velocity $v(y)$ at position y is bounded by some function $z(y)$, such that $0 < v(y) \leq z(y)$. Then the time t to reach the point x starting

from the origin satisfies the bound

$$t = \int_0^x \frac{dy}{v(y)} \geq \int_0^x \frac{dy}{z(y)}. \quad (159)$$

Now let us consider an algorithm a , its cost being $C(a)$ and denote by $m_1, \dots, m_{C(a)}$ the steps at which the RW has been probed by the searcher – which has eventually found the maximum M_n after $C(a)$ probes. To be sure that the maximum is not in the interval $[m_i, m_{i+1}]$, the potential maximum of the RW between these two steps, which is $(X_{m_i} + X_{m_{i+1}} + m_{i+1} - m_i)/2$ (see Fig. 11), must be smaller than M_n (by definition of the maximum). Hence this yields the following inequality

$$m_{i+1} - m_i \leq 2M_n - X_{m_i} - X_{m_{i+1}}. \quad (160)$$

Notice that $m_{i+1} - m_i$ can be seen as the velocity $v(m_i)$ of the algorithm at point m_i . One can further argue [30], using the fact most of the RWs are “slowly varying”, that $2M_n - X_{m_i} - X_{m_{i+1}} \sim 2(M_n - X_{m_i})$ when n is large. Hence

$$Z_k = 2(M_n - X_k) \quad (161)$$

can be viewed as the speed limit at step k of the random walk. Finally, by analogy with (159), $C(a)$ satisfies

$$C(a) \geq \sum_{k=1}^n \frac{1}{Z_k} = \frac{1}{2} \sum_{k=1}^n \frac{1}{M_n - X_k}, \quad (162)$$

which in the continuum limit yields the functional of the maximum I in Eq. (4). It is rather clear that these heuristic arguments leading to Eq. (162) can be straightforwardly extended to the case of the Random Walk bridge, $X_{i,B}$, which is a RW conditioned to start and end at the origin $X_{0,B} = X_{n,B} = 0$. Of course in this case the maximum M_n in (162) is then replaced by the maximum of the Brownian bridge $M_{n,B} = \max_{1 \leq i \leq n} X_{i,B}$.

C.2 Description of the Odlyzko’s algorithm

Here we describe in more detail Odlyzko’s algorithm which finds the maximum of a random walk $X_{i+1} = X_i \pm 1$ with equal probability $1/2$ (starting from $X_0 = 0$). Let c be a positive real number, which is sufficiently large. The algorithm is essentially based on the fact that most of the RWs has “slow variations” (SV), i.e., check the identity [30]:

$$|X_{i+k} - X_i| \leq c\sqrt{k \log n}, \quad \forall i, k \text{ with } i+k \leq n. \quad (163)$$

Indeed if c is large enough, the probability that a realization of the RW does not satisfy the SV property (163) decays as n^{-1} . This statement can be easily shown, as in [30, 33], by using that for fixed j , $\Pr(|X_j| > x) \leq 2\exp(-x^2/(2j))$ [the so called Chernoff’s bound, see [49] p. 12]. Although the realizations of the RW that do not satisfy (163) necessitates a large number of probes $\sim n$, their contribution to the average cost of the algorithm turns out to be negligible as they occur with a very small probability $\propto 1/n$. On the other hand, as we shall see below, it is relatively easy to find the maximum of a RW which satisfies the “SV” property.

The algorithm proposed by Odlyzko consists in two steps:

- In a first stage, one searches a good estimate M^* of M_n . This is done by probing $X_N, X_{2N}, X_{3N}, \dots$ where $N = \lfloor \sqrt{n \log n} \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer not larger than x . If the algorithm finds, here or later, a violation of the SV inequality (163), one has to probe all the positions of the RW (but this happens very rarely). We denote by $M' = \max\{X_0, X_N, X_{2N}, X_{3N}, \dots\} \leq M_n$. If the RW satisfies SV (163), then

$$M_n - M' \leq c\sqrt{N \log n} = cn^{1/4} \log n. \quad (164)$$

Indeed, if we denote by k_{\max} such that $X_{k_{\max}N} \leq M_n \leq X_{(k_{\max}+1)N}$ then $M_n - \max(X_{k_{\max}N}, X_{(k_{\max}+1)N}) \leq c\sqrt{N \log n}$, which follows from (163), and which implies (164) as $M' \geq \max(X_{k_{\max}N}, X_{(k_{\max}+1)N})$. As we discuss it below, it turns out that this estimate M' of M_n (164) is however not precise enough for the forthcoming steps of the algorithm. It is indeed necessary to scan the neighborhood of the large X_{rN} ’s on a finer window. If for some integer r one finds

$$X_{rN} \geq M' - cn^{1/4} \log n, \quad (165)$$

we probe $X_{rN \pm jK}$, $j = 1, 2, \dots, \lfloor N/K \rfloor$, $K = \lfloor n^{1/4} \rfloor$. If the RW has SV, then any k with $X_k = M_n$ must be as close as of a rN for some r for which (165) is true. We now denote M^* the maximum of all probes found until now. Because we scan with intervals $\leq n^{1/4} \log n$ around the maximum, the SV inequality (163) give

$$0 \leq M_n - M^* \leq c\sqrt{n^{1/4} \log^2 n} \leq n^{1/6}. \quad (166)$$

One can prove [30] that the average cost of this first phase of the algorithm is of order $\mathcal{O}(\sqrt{n}/\log n)$ negligible compared to the cost of the second phase, that we now describe, and which is of order $\mathcal{O}(\sqrt{n})$.

• With this estimate M^* of the actual maximum M_n , the second phase will eventually find M_n in a number of probes that is of order $\mathcal{O}(\sqrt{n})$, which is the leading contribution to the cost of this algorithm. To do this, we will scan the sample path from left to right as follows. We introduce m the index of the RW position X_m which is currently probed by the algorithm. We start with $m = 0$ and we denote by $M^\#$ the greatest position probed so far by the algorithm including M^* . At each step of this phase, two cases may occur:

- (i) If $M^\# - X_m \leq n^{1/6}$, the algorithm will probe the right neighbor of X_m and m is incremented by 1, $m \rightarrow m + 1$.
- (ii) If $M^\# - X_m > n^{1/6}$, this means that the algorithm is still far from the maximum, because we know that $M_n - M^\# \leq n^{1/6}$. In this case, the immediate vicinity of X_m does not need to be explored and the strategy is to jump from X_m to X_{m+k} , where k is still to be determined. In order to be sure that the RW does not exceed $M^\#$ between m and $m + k$, we must have in mind the upper envelope of the RW on that interval $[m, m + k]$ (see Fig. 11). Hence we impose the following bound

$$k \leq 2(M^\# - X_m) + (X_m - X_{m+k}). \quad (167)$$

The first term in the right hand side of this inequality (167), $2(M^\# - X_m)$, is larger than $2n^{1/6}$, while the second term, $(X_m - X_{m+k})$ is bounded by $c\sqrt{k \log n}$, thanks to SV (163) – as stated above, if $X_{m+k} - X_m$ does not satisfy the SV inequality (163), we abort this approach and probe every position. Hence we can choose k slightly smaller than $2(M^\# - X_m)$. If $m + k > n$, we probe X_n and stop. When the full path has been scanned, the maximum M_n of the RW has been found by the algorithm.

For a RW which satisfies SV (163), one can show [30] that the major contribution to the cost of the algorithm is when $M^\# - X_m > n^{1/6}$. Indeed, one can show that the contributions of the probes of the type (i) to the cost of the algorithm is of the order $\mathcal{O}(n^{1/3})$. In fact, one can show that if the estimate M^* of M_n is such that $M_n - M^* < n^\alpha$ then the cost of these contributions is of order $\mathcal{O}(n^{2\alpha})$. If we want that the cost of this part of the algorithm to be smaller than the cost of the last one, which is of order $\mathcal{O}(\sqrt{n})$, then this requires $2\alpha < 1/2$, for instance $2\alpha = 1/3$, hence the choice $\alpha = 1/6$ made by Odlyzko [30].

The step size k is slightly smaller than $2(M_n - X_m)$ and we need only one probe to control the k positions between m and $m + k$. Since k can be interpreted as the velocity of the algorithm [see Eq. (159)], the average cost of the algorithm is, at leading order when n goes to infinity, $\langle C(\text{Od}) \rangle$ given by

$$\langle C(\text{Od}) \rangle = \frac{1}{2} \left\langle \sum_{i=0}^n \frac{1}{M_n - X_i + 1} \right\rangle, \quad (168)$$

where we recall that $\langle \dots \rangle$ denotes an average over the different realizations of the RW X_i 's. When n goes to infinity, the RW becomes a BM and

$$\frac{C(\text{Od})}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} I = \frac{1}{2} \int_0^1 \frac{d\tau}{x_{\max} - x(\tau)}, \quad (169)$$

as described in the text in (4).

C.3 Odlyzko's algorithm for the Bridge

It is easy to check that the arguments presented above can be easily transposed to the case of a random walk bridge. In particular, given that the bridge is pinned at both extremities $X_{0,B} = X_{n,B} = 0$, its variations are typically smaller than the one of the free walk and hence the properties of "slow variations" (163), which plays a crucial role in this algorithm, would follow naturally. Therefore we conjecture that Odlyzko's algorithm would be the optimal one to find the maximum $M_{n,B}$ and its cost would be given by $(1/2)T_{\alpha=-1}^B(t)$ given in Eq. (13).

D Useful integrals involving confluent hypergeometric functions relevant for the case $V(x) = 1/x$

D.1 An integral involving a single confluent hypergeometric function

For the analysis of the functional $T_{\alpha=-1}(t)$ [see Eq. (127)], a useful integral involving the confluent hypergeometric function $U(a, 2, z)$ is the following (see [50] as well as Mathematica):

$$\tilde{\varphi}(s) = 2^{3/2} \lambda \Gamma(\lambda/\sqrt{2s}) \int_0^\infty e^{-\sqrt{2s}y} y U\left(1 + \frac{\lambda}{\sqrt{2s}}, 2, 2\sqrt{2s}y\right) dy \quad (170)$$

$$= \frac{1}{\sqrt{2s}} \left(\sqrt{2s} - 2\pi \lambda \csc\left(\frac{\pi\lambda}{\sqrt{2s}}\right) - \lambda H\left(-\frac{1}{2} - \frac{\lambda}{2\sqrt{2s}}\right) + \lambda H\left(-\frac{\lambda}{2\sqrt{2s}}\right) \right), \quad (171)$$

where $\csc(x) = 1/\sin x$ and $H(x)$ are harmonic numbers, $H(x) = \psi(x) + \gamma_E$ where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the di-gamma function and γ_E the Euler constant. The function $H(x)$ admits the following series expansion

$$H(x) = \sum_{j=0}^{\infty} (-1)^j \zeta(j+2) x^{j+2}, \quad (172)$$

where $\zeta(x)$ is the Riemann zeta function. By combining (170), together with (172) one arrives straightforwardly to the formula given in Eq. (128) in the main text.

D.2 An integral involving the product of two confluent hypergeometric functions

To compute the amplitudes c_E such that the functions $\phi_E(x)$ in (138) with $d_E = 0$ satisfy the orthogonality condition in Eq. (139) we used the following relation, derived by Landau and Lifshitz [51] (see formula (f.9) in Appendix f):

$$\begin{aligned} J &= \int_0^\infty e^{-\lambda z} z^{\gamma-1} {}_1F_1(\alpha, \gamma, kz) {}_1F_1(\alpha', \gamma', k'z) dz \\ &= \Gamma(\gamma) \lambda^{\alpha+\alpha'-\gamma} (\lambda-k)^{-\alpha} (\lambda-k')^{-\alpha'} {}_2F_1\left(\alpha, \alpha', \gamma, \frac{kk'}{(\lambda-k)(\lambda-k')}\right), \end{aligned} \quad (173)$$

where ${}_2F_1(\alpha, \alpha', \gamma, z)$ is a generalized hypergeometric series. Such integrals (173) arise naturally in the study of certain matrix elements of quantum Hamiltonian involving Coulomb interactions. In our case (138), one has $\alpha = 1 - is/\sqrt{E}$, $\alpha' = 1 - is/\sqrt{E'}$, $\gamma = \gamma' = 2$, $k = 2i\sqrt{2E}$, $k' = 2i\sqrt{2E'}$ and $\lambda = 2\sqrt{2s}$. Hence the desired formula in our case (139) can be obtained by differentiating (173) once wrt λ and analyzing in detail the limit $k \rightarrow k'$ of the resulting formula (173). These somewhat cumbersome manipulations yield the expression for c_E given in (140).

E Numerical simulations of constrained Brownian motion

In this appendix, we describe the algorithms that we have used here to simulate various constrained Brownian motions. We refer the interested reader to [52] for an extended discussion of these algorithms.

E.1 Brownian motion

In order to simulate a Brownian motion $x(\tau)$, we consider the discrete random walk

$$\begin{cases} X_0 = 0 \\ X_i = X_{i-1} + \frac{\eta_i}{\sqrt{N}}, \quad i \in [1, N] \end{cases} \quad (174)$$

where η_i 's are identical and independent Gaussian standard variables of variance unity. When N goes to infinity, $X_{[\tau N]} \rightarrow x(\tau)$ where $x(\tau)$ is a Brownian motion, with $\tau \in [0, 1]$: $\dot{x}(\tau) = \zeta(\tau)$, where $\zeta(\tau)$ is a Gaussian white noise $\langle \zeta(\tau)\zeta(\tau') \rangle = \delta(\tau - \tau')$. This is the building block (174), to simulate different constrained Brownian motions.

```
//generation of Brownian Motion
void BM ( int N, double *X, gsl_rng * r)
{
  //r is the 'seeds' of the random number generator.
  int i;
  X[0]=0;
  for(i=1; i<N; i++)
  {
    X[i]=X[i-1]+gsl_ran_gaussian (r, 1)/sqrt((double)N);
  }
}
```

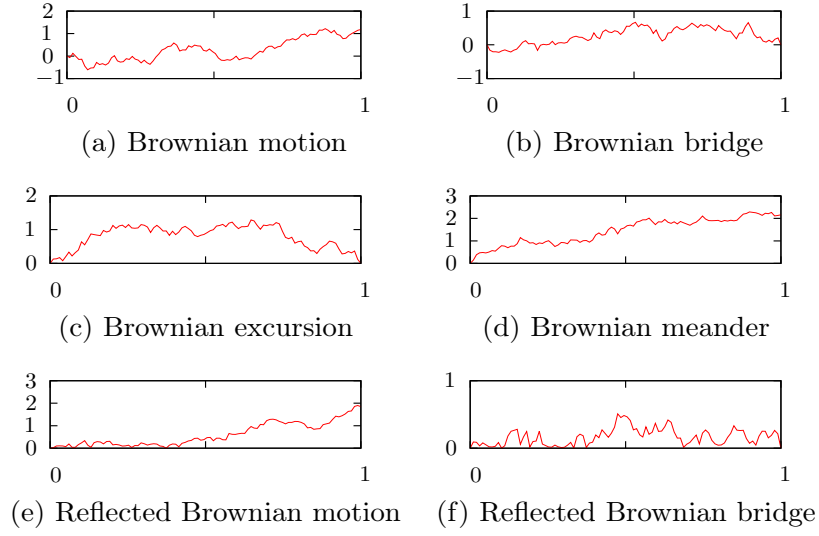


Fig. 13 Example of different constrained Brownian motions studied in the present paper.

E.2 Brownian bridge

For a Brownian bridge $x_B(\tau)$, which is a Brownian motion starting and ending at the origin $x_B(0) = x_B(1) = 0$, we use the identity $x(\tau) - \tau x(1) = x_B(\tau)$

$$Y_i = X_i - \frac{i}{N} X_N, \quad i \in [0, N], \quad (175)$$

where X_i 's are generated by (174). One can show that $Y_{[\tau N]}$ converges to a Brownian bridge $x_B(\tau)$.

```
//generation of Brownian Bridge
void BB ( int N, double *X, gsl_rng * r)
{
    BM(N,X,r);
    int i;
    for(i=1; i<N; i++)
    {
        X[i]=X[i]-(double)i/(N-1)*X[N-1];
    }
}
```

E.3 Brownian excursion

For a Brownian excursion $x_E(\tau)$, which is a Brownian motion that starts and ends at the origin $x_E(0) = x_E(1) = 0$ and staying positive in the interval $[0, 1]$, we use the identity $\sqrt{[x_{B,1}(\tau)]^2 + [x_{B,2}(\tau)]^2 + [x_{B,3}(\tau)]^2} = x_E(\tau)$ where $x_{B,1}, x_{B,2}$ and $x_{B,3}$ are three independent Brownian bridges [53, 54]. Hence we simulate

$$E_i = \sqrt{Y_{1,i}^2 + Y_{2,i}^2 + Y_{3,i}^2}, \quad i \in [0, N] \quad (176)$$

where $Y_{1,i}, Y_{2,i}$ and $Y_{3,i}$ are three independent realisations of (175). $E_{[\tau N]}$ converges to a Brownian excursion $x_E(\tau)$.

```
//generation of Brownian Excursion
void BE ( int N, double *X, gsl_rng * r)
{
    double X1[N];
    double X2[N];
    double X3[N];
```

```

BB(N,X1,r);
BB(N,X2,r);
BB(N,X3,r);
int i;
for(i=1; i<N; i++)
{
    X[i]=sqrt(X1[i]*X1[i]+X2[i]*X2[i]+X3[i]*X3[i]);
}
}

```

E.4 Brownian meander

For a Brownian meander $x_{Me}(\tau)$, a Brownian motion which begins at the origin and stays positive on $[0, 1]$, one can show that the PDF of its final position $x_F > 0$ at time 1 is $p(x_F) = x_F e^{-x_F^2/2}$. One can then use the following representation of the meander ending at x_F : [53, 54] $\sqrt{[x_{B,1}(\tau)]^2 + [x_{B,2}(\tau)]^2 + [x_{B,3}(\tau) + \tau x_F]^2} = x_{Me}(\tau)$ where $x_{B,1}, x_{B,2}$ and $x_{B,3}$ are three independent Brownian bridges and x_F is a random variable drawn from $p(x_F) = x_F e^{-x_F^2/2}$. Hence the Brownian meander $x_{Me}(\tau)$ can be generated numerically as

$$M_i = \sqrt{Y_{1,i}^2 + Y_{2,i}^2 + \left(Y_{3,i} + f \frac{i}{N}\right)^2}, \quad i \in [0, N] \quad (177)$$

where Y_1, Y_2 and Y_3 are three independent realizations of (175), where $f > 0$ is a random variable, whose PDF is given by $p(f) = f e^{-f^2/2}$. $M_{[tN]}$ converges to a Brownian meander $x_{Me}(\tau)$.

```

//generation of Brownian Meander
void BMe ( int N, double *X, gsl_rng * r)
{
    double f=gsl_ran_rayleigh (r,1);

    double X1[N];
    double X2[N];
    double X3[N];
    BB(N,X1,r);
    BB(N,X2,r);
    BB(N,X3,r);
    int i;
    for(i=0; i<N; i++)
    {
        X[i]=sqrt((X1[i]+f*(double)i/(N-1))*(X1[i]+f*(double)i/(N-1))+X2[i]*X2[i]+X3[i]*X3[i]);
    }
}

```


References

1. S. Chandrasekhar, *Stochastic Problems in Physics and Astronomy*, Rev. Mod. Phys. **15**, 1 (1943).
2. W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York (1968).
3. B. Hughes, *Random walks and random environments*, Clarendon Press, Oxford, (1968).
4. D. E. Koshland, *Bacterial Chemotaxis as a Model Behavioral System*, Raven, New York (1980).
5. S. Asmussen, *Applied Probability and Queues*, Springer, New York (2003).
6. M. J. Kearney, *On a random area variable arising in discrete-time queues and compact directed percolation*, J. Phys. A **37**, 8421 (2004).
7. M. J. Kearney, S. N. Majumdar, *On the area under a continuous time Brownian motion till its first-passage time*, J. Phys. A: Math. Gen. **38**, 4097 (2005).
8. S. N. Majumdar, *Brownian Functionals in Physics and Computer Science*, Curr. Sci. **89**, 2076 (2005).
9. S. N. Majumdar, *Universal first-passage properties of discrete-time random walks and Lévy flights on a line: Statistics of the global maximum and records*, Physica A **389**, 4299 (2010).
10. R. J. Williams, *Introduction to the Mathematics of Finance*, AMS, Providence (2006).
11. S. N. Majumdar, J. P. Bouchaud, *Optimal time to sell a stock in the Black-Scholes model: comment on 'Thou shalt buy and hold'*, by A. Shiryaev, Z. Xu and XY Zhou, Quant. Fin., **8**, 753 (2008).
12. A. Comtet, J. Desbois, C. Texier, *Functionals of Brownian motion, localization and metric graphs*, J. Phys. A **38**, R341 (2005).
13. M. Yor, *Exponential Functionals of Brownian Motion and Related Topics*, Springer, Berlin (2000).
14. J. Pitman, *The distribution of local times of Brownian bridge*, Lecture Notes in Math., vol. 1709, pp. 388-394, Springer, Berlin (1999).
15. D. A. Darling, *On the supremum of certain Gaussian processes*, Ann. Probab. **11**, 803 (1983).
16. G. Louchard, *Kac's formula, Levy's local time and Brownian excursion*, J. Appl. Prob. **21**, 479 (1984).
17. P. Flajolet, P. Poblete, A. Viola, *On the analysis of linear probing hashing*, Algorithmica **22**, 490 (1998).
18. S. Janson, G. Louchard, *Tail estimates for the Brownian excursion area and other Brownian areas*, Electronic J. Probab. **12**, 1600 (2007).
19. S. N. Majumdar, A. Comtet, *Exact maximal height distribution of fluctuating interfaces*, Phys. Rev. Lett. **92**, 225501 (2004).
20. S. N. Majumdar, A. Comtet, *Airy distribution function: from the area under a Brownian excursion to the maximal height of fluctuating interfaces*, J. Stat. Phys. **119**, 777 (2005).
21. D. A. Kessler, S. Medalion, E. Barkai, *The Distribution of the Area Under a Bessel Excursion and its Moments*, J. Stat. Phys. **156**, 686 (2014).
22. F. Black, M. Scholes, *The pricing of options and corporate liabilities*, J. Pol. Econ. **81**, 637 (1973).
23. H. Kesten, M. V. Kozlov, F. Spitzer, *A limit law for random walk in a random environment*, Compos. Math. **30**, 145 (1975).
24. G. Oshanin, A. Mogutov, M. Moreau, *Steady flux in a continuous-space Sinai chain*, J. Stat. Phys. **73** (1993).
25. C. Monthus, A. Comtet, *On the flux distribution in a one dimensional disordered system*, J. Phys. I (France) **4**, 635 (1994).
26. G. Oshanin, A. Rosso, G. Schehr, *Anomalous fluctuations of currents in Sinai-type random chains with strongly correlated disorder*, Phys. Rev. Lett. **110**, 100602 (2013).
27. M. Kac, *On distributions of certain Wiener functionals*, Trans. Am. Math. Soc. **65**, 1 (1949).
28. S. Sabhapandit, S. N. Majumdar, *Density of near-extreme events*, Phys. Rev. Lett. **98**, 140201 (2007).
29. A. Perret, A. Comtet, S. N. Majumdar, G. Schehr, *Near-extreme statistics of Brownian motion*, Phys. Rev. Lett. **111**, 240601 (2013).
30. A. M. Odlyzko, *Search for the maximum of a random walk*, Random Struct. Algor. **6**, 275 (1995).
31. H. K. Hwang, *A constant arising from the analysis of algorithms for determining the maximum of a random walk*, Random Struct. Algor. **10**, 333 (1997).
32. P. Chassaing, *How many probes are needed to compute the maximum of a random walk?*, Stoch. Proc. Appl. **81**, 129 (1999).
33. P. Chassaing, J. F. Marckert, M. Yor, *A stochastically quasi-optimal search algorithm for the maximum of the simple random walk*, Ann. Appl. Probab. **13**, 1264 (2003).
34. P. Biane, M. Yor, *Valeurs principales associées aux temps locaux browniens*, Bull. Sci. Maths **111**, 23 (1987).
35. P. Chassaing, J.F. Marckert, M. Yor, *The height and width of simple trees*, Math. Computer Science, pp. 17-30, Birkhäuser Basel (2000).
36. W. Vervaat, *A relation between Brownian bridge and Brownian excursion*, Ann. Probab. **7**, 143 (1979).
37. L. Takács, *A Bernoulli excursion and its various applications*, Adv. Appl. Prob. **23**, 557 (1991).
38. L. Takács, *Limit distributions for the Bernoulli meander*, J. Appl. Prob. **32**, 375 (1995).
39. L. Takács, *Brownian local times*, J. Appl. Math. Stoch. Anal. **8**, 209 (1995).
40. T. W. Burkhardt, G. Gyögyi, N. R. Moloney, Z. Racz, *Extreme statistics for time series: Distribution of the maximum relative to the initial value*, Phys. Rev. E, **76**(4), 041119 (2007).
41. P. Lévy, *Sur certains processus stochastiques homogènes*, Compos. Math. **7**, 283 (1940).
42. H. Krivine, *Exercices de mathématiques pour physiciens, corrigés et commentés*, Cassini, Paris (2003).
43. W. Feller, *The asymptotic distribution of the range of sums of independent random variables*, Ann. Math. Stat. **22**, 427 (1951).
44. A. Kundu, S. N. Majumdar, G. Schehr, *Exact distributions of the number of distinct and common sites visited by N independent random walkers*, Phys. Rev. Lett. **110**, 220602 (2013).

-
45. K. L. Chung, *Excursions in Brownian motion*, Ark. Mat. **14**(2), 155 (1976).
 46. L. Takács, *Limit theorems for random trees*, Proc. Natl. Acad. Sci. USA **89**(11), 5011 (1992).
 47. G. Schehr, S. N. Majumdar, A. Comtet, J. Randon-Furling, *Exact distribution of the maximal height of p vicious walkers*, Phys. Rev. Lett. **101**, 150601 (2008).
 48. P. Chassaing, G. Louchard, *Reflected Brownian bridge area conditioned on its local time at the origin*, J. Algorithm **44**(1), 29 (2002).
 49. B. Bollobás, *Random Graphs*, Academic Press, Boston (1985).
 50. I. S. Gradshteyn, I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic Press (2000).
 51. L. D. Landau, E. M. Lifshitz, *Quantum mechanics: non-relativistic theory*, Pergamon, London (1981).
 52. L. Devroye, *On exact simulation algorithms for some distributions related to Brownian motion and Brownian meanders*, Recent Developments in App. Prob. and Stat., Springer, 1 (2010).
 53. D. Williams, *Decomposing the Brownian path*, B. Am. Math. Soc. **76**, 871 (1970).
 54. J. P. Imhof, *Density factorizations for Brownian motion, meander and the three-dimensional Bessel process, and applications*, J. Appl. Probab. **21**, 500 (1984).

Finite N corrections to the limiting distribution of the smallest eigenvalue of Wishart complex matrices.

[128] A. Perret, G. Schehr (2015).

Prépublication arXiv :1506.02387

Soumis à Random Matrices : Theory and Applications

Finite N corrections to the limiting distribution of the smallest eigenvalue of Wishart complex matrices

Anthony Perret

*Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100,
91405 Orsay Cedex, France **

Grégory Schehr

*Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bât. 100,
91405 Orsay Cedex, France †*

We study the probability distribution function (PDF) of the smallest eigenvalue of Laguerre-Wishart matrices $W = X^\dagger X$ where X is a random $M \times N$ ($M \geq N$) matrix, with complex Gaussian independent entries. We compute this PDF in terms of semi-classical orthogonal polynomials, which are deformations of Laguerre polynomials. By analyzing these polynomials, and their associated recurrence relations, in the limit of large N , large M with $M/N \rightarrow 1$ – i.e. for quasi-square large matrices X – we show that this PDF, in the hard edge limit, can be expressed in terms of the solution of a Painlevé III equation, as found by Tracy and Widom, using Fredholm operators techniques. Furthermore, our method allows us to compute explicitly the first $1/N$ corrections to this limiting distribution at the hard edge. Our computations confirm a recent conjecture by Edelman, Guionnet and Pécché. We also study the soft edge limit, when $M - N \sim \mathcal{O}(N)$, for which we conjecture the form of the first correction to the limiting distribution of the smallest eigenvalue.

1. Introduction

The study of extreme eigenvalue statistics in Random Matrix Theory (RMT) has attracted much attention during the last twenty years. In particular, the Tracy-Widom (TW) distributions [1,2] describing the largest eigenvalue λ_{\max} (as well as the smallest one, λ_{\min}) in the classical Gaussian ensembles, orthogonal (GOE, $\beta = 1$), unitary (GUE, $\beta = 2$) and symplectic (GSE, $\beta = 4$) – where β is the Dyson index – have become cornerstones of the theory of extreme value statistics of strongly correlated variables. Quite remarkably, it was shown that the TW distributions, denoted by \mathcal{F}_β in the following, appear in a wide variety of problems [3], a priori not directly related to RMT, ranging from the longest increasing sequence of random permutations of integers [4], stochastic growth and related directed polymer models in the Kardar-Parisi-Zhang [5] universality class [6,7,8,9,10,11] and

*anthony.perret@u-psud.fr

†gregory.schehr@u-psud.fr

sequence alignment problems [12] to non-intersecting interfaces and Brownian motions [13,14,15] as well as in finance [16]. While the TW distribution describes the typical fluctuations of λ_{\max} and λ_{\min} , a large body of work has also been devoted to the study of large deviations of extreme eigenvalues in Gaussian ensembles [17].

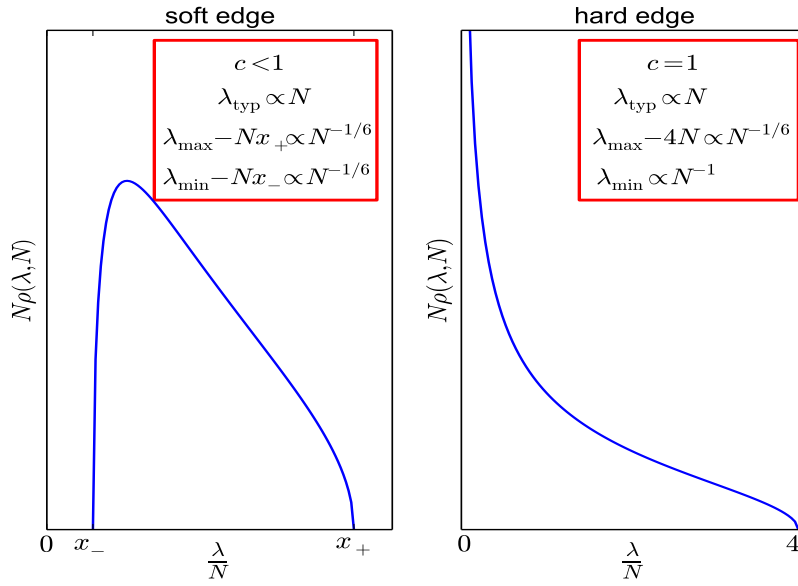


Figure 1. Plot of the Marchenko-Pastur distribution for $c < 1$ (**left**) and for $c = 1$ (**right**).

Another interesting ensemble of random matrices, which we focus on in this paper, is the so-called Wishart-Laguerre ensemble – here we focus on the case of complex matrices ($\beta = 2$). Let X be a $M \times N$ rectangular matrix with i.i.d. complex Gaussian entries and $M - N = a \geq 0$. The Wishart-Laguerre matrix W is defined as $W = X^\dagger X$ which is thus a $N \times N$ Hermitian matrix, having N real and positive eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$. The joint probability density function (PDF) of these N eigenvalues is given by [18,19]

$$P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i < j}^N (\lambda_i - \lambda_j)^2 \exp \left(- \sum_{i=1}^N \lambda_i \right) \prod_{i=1}^N \lambda_i^a, \quad (1.1)$$

where $\lambda_1, \lambda_2, \dots, \lambda_N$ are positive and where Z_N is a normalization constant, depending on a . These Wishart-Laguerre matrices play an important role in statistics, in particular in principal component analysis, where the matrix W is a covariance matrix. Hence in this case both M and N , and thus $a = M - N$, are positive integers. However, this joint PDF in Eq. (1.1) is well defined for any real value of $a \geq 0$,

which, for some non-integer value of a , may have some physical applications. An interesting example is the case of non-intersecting Brownian excursions [20], i.e., N non-colliding positive Brownian paths $x_i(t) \geq 0$ on the unit time interval $t \in [0, 1]$ constrained to start and end at the origin $x_i(0) = x_i(1) = 0$. The joint PDF of the positions of the N walkers, at a given time t , can indeed be written as [21]

$$P_{\text{joint}}(x_1, \dots, x_N; t) = \frac{1}{z_N(t)} \prod_{i=1}^N x_i^2 \prod_{i < j}^N (x_i^2 - x_j^2)^2 \exp \left(- \sum_{i=1}^N \frac{x_i^2}{\sigma^2(t)} \right), \quad (1.2)$$

where $\sigma^2(t) = 2t(1-t)$ and $z_N(t)$ is a normalization constant. From this expression (1.2) we obtain that the scaled variables $x_i^2/\sigma^2(t)$ behave statistically like the eigenvalues of random matrices from the Wishart-Laguerre ensemble (1.1) with a non-integer parameter $a = 1/2$. Hence it is physically relevant to study the joint distribution in Eq. (1.1) for any real $a \geq 0$.

A first important characteristic associated to this ensemble (1.1) is the mean density of eigenvalues, $\rho(\lambda, N)$, defined by

$$\rho(\lambda, N) = \frac{1}{N} \sum_{i=1}^N \overline{\delta(\lambda_i - \lambda)}, \quad (1.3)$$

where the overline denotes an average over the different realizations of the random variables λ_i 's according to the joint PDF in Eq. (1.1). In the large N limit, it is well known that $\rho(\lambda, N)$ is given by the Marchenko-Pastur (MP) distribution (see Fig. 1)

$$\rho(\lambda, N) \xrightarrow{N \rightarrow \infty} \frac{1}{N} \rho_{\text{MP}} \left(\frac{\lambda}{N} \right), \quad (1.4)$$

$$\rho_{\text{MP}}(x) = \frac{\sqrt{(x - x_-)(x_+ - x)}}{2\pi x}, \quad (1.5)$$

where $x_{\pm} = (c^{-1/2} \pm 1)^2$ are the right and left edges of the support, with $c = N/M \leq 1$. The case where $c < 1$ corresponds to the case where $a \sim \mathcal{O}(N)$. Here, we mainly focus on the case $c = 1$ which corresponds instead to the case where a is finite, while both N and M are large. In this case, which we will mainly focus on in this paper, the MP distribution takes the particular form (see the left panel of Fig. 1)

$$\rho_{\text{MP}}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}}. \quad (1.6)$$

At the right edge, near $x_+ = 4$, $\rho_{\text{MP}}(x)$ vanishes as a square-root, $\rho_{\text{MP}}(x) \propto \sqrt{4-x}$ and therefore the fluctuations near this *soft* edge are governed, for large N , by the Airy kernel. In particular, the distribution of the largest eigenvalue λ_{max} , appropriately shifted and scaled, converges, when $N \rightarrow \infty$, to the TW distribution \mathcal{F}_2 mentioned above (the same as for GUE) [6,22]. It is now well known that this distribution can be expressed in terms of a special solution of a Painlevé II equation. While this connection to Painlevé transcendents was initially obtained by Tracy and

Widom using Fredholm operator techniques, Nadal and Majumdar [23] provided, more recently, a derivation of this result (for $\beta = 2$) using semi-classical orthogonal polynomials (OPs), see also Ref. [24]. This method is at the heart of the present paper.

On the other hand, at the left edge, $x_- = 0$, $\rho_{\text{MP}}(x)$ has a square-root singularity, $\rho_{\text{MP}}(x) \propto 1/\sqrt{x}$ (see the left panel of Fig. 1). What about the fluctuations of the smallest eigenvalue λ_{\min} in this case? One can estimate the typical scale of λ_{\min} , for large N , by considering that there is typically one eigenvalue in the interval $[0, \lambda_{\min}]$, i.e.,

$$N \int_0^{\lambda_{\min}} \rho(\lambda, N) d\lambda \sim \mathcal{O}(1), \quad (1.7)$$

which implies that $\lambda_{\min} \sim \mathcal{O}(1/N)$. Within this scale $\mathcal{O}(1/N)$, the fluctuations are governed by the Bessel kernel [25,26]. Furthermore, it has been shown that the distribution of $N\lambda_{\min}$ converges to a limiting form which (i) is different from the standard TW distribution \mathcal{F}_2 and depends continuously on the exponent a in (1.1) and (ii) can be written in terms of a special solution of a Painlevé III equation [27]. If one introduces $F_N(t) = \text{Pr}(\lambda_{\min} \geq t)$ then, one has indeed

$$\lim_{N \rightarrow \infty} F_N\left(\frac{x}{N}\right) = F_{\infty}(x), \quad F_{\infty}(x) = \exp\left(\int_0^x \frac{f(u)}{u} du\right), \quad (1.8)$$

where $f(x)$ is the unique solution of a Painlevé III [27]:

$$(xf'')^2 + 4f'(1+f')(xf' - f) = (af')^2, \quad (1.9)$$

satisfying

$$f(x) \sim -\frac{x^{a+1}}{\Gamma(a+1)\Gamma(a+2)}, \quad \text{as } x \rightarrow 0. \quad (1.10)$$

This result (1.8) was shown by Tracy and Widom using Fredholm operator techniques [27]. Note that for integer values of a , the limiting distribution $F_{\infty}(x)$ can be written as an $a \times a$ determinant whose entries are expressed in terms of Bessel functions (see Eq. (C.1) below) – a result which can be obtained by clever manipulations of determinants [28]. In particular for $a = 0$ the result is extremely simple as $F_N(x) = \exp(-Nx)$ for all N , implying $f(x) = -x$, which is obviously solution of Eq. (1.9) with the boundary condition (1.10).

The limiting distribution of λ_{\min} for complex Wishart matrices, in the limit $N \rightarrow \infty$, is thus well known (1.8) – we refer the reader to Ref. [29] to a recent work on the smallest eigenvalue for real Wishart matrices in the hard edge limit. What about the finite N corrections to this asymptotic form? Such a question is quite natural for practical applications of extreme value statistics (EVS), where one always deals with finite samples – here matrices of finite size. This issue was recently revisited for EVS of independent and identically distributed random variables using a renormalization group approach [30]. For EVS of strongly correlated variables, there are actually few cases where these corrections have been worked out

explicitly, including random walks [31], the *largest* eigenvalue of random matrices belonging to various ensembles [32,33,34,35], non-intersecting Brownian motions [36] or (Poissonized) random matchings [37]. For *real* Wishart matrices, the first corrections to the limiting distribution of the *smallest* eigenvalue in the soft edge limit were studied in Ref. [35] where it was shown that corrections to the limiting distribution of λ_{\min} and λ_{\max} are quite different, although the limiting distributions for both observables are actually the same, namely the TW distribution for GOE, \mathcal{F}_1 . What about the corrections to the limiting distribution F_∞ in Eq. (1.8) of λ_{\min} for complex Wishart matrices in the hard edge limit? This question was recently raised by Edelman, Guionnet et P      [38] in their study of finite size covariance matrices with non-Gaussian entries. Based on the large N expansion of the exact formulas obtained in Ref. [28] for small integer values of a , they conjectured the following form of the first $1/N$ -correction

$$F_N\left(\frac{x}{N}\right) = F_\infty(x) + \frac{a}{2N}x F'_\infty(x) + o\left(\frac{1}{N}\right). \quad (1.11)$$

Note that this first $1/N$ correction in Eq. (1.11) can be interpreted as a correction to the width, i.e.,

$$F_N\left(\frac{x}{N}\right) = F_\infty\left(x\left(1 + \frac{a}{2N}\right)\right) + o\left(\frac{1}{N}\right). \quad (1.12)$$

It is interesting to notice [37] that for most of the cases which have been studied in RMT [32,33,34], it was actually found that the first order correction to the limiting distribution of extreme eigenvalue actually corresponds to a correction of the scaling variable, as in Eq. (1.12). One exception concerns the smallest eigenvalue of real Wishart matrices in the soft-edge limit, where the first correction has a more complicated structure [35].

The main goal of this paper is to provide an explicit computation of this first correction in the hard edge limit and we will show that it has indeed the conjectured form given above in Eq. (1.11). To perform this computation, we will use a method relying on semi-classical OPs, in the spirit of Refs. [23] and [24,39]. As we will see, our method does not only allow us to compute explicitly the first $1/N$ corrections but provides also a rather straightforward derivation of the expression for the limiting distribution $F_\infty(x)$ in terms of the solution of a Painlev   III equation, without using Fredholm operators theory but relying instead only on the recurrence relations associated to the (semi-classical) OPs system. Finally, we will also study the first finite N corrections to the limiting distribution of λ_{\min} at the soft edge.

Note that after the results obtained in the present paper were presented in a conference [40], another independent proof of the conjecture in Eq. (1.11) was achieved in Ref. [41], using operator theoretic techniques. More recently, yet another independent proof of this conjecture was given in Ref. [42].

2. Summary of main results and outline of the paper

The distribution of the smallest eigenvalue $\lambda_{\min} = \min_{1 \leq i \leq N} \lambda_i$ is given by

$$F_N(t) = \text{Prob}(\lambda_{\min} \geq t) = \int_t^\infty d\lambda_1 \int_t^\infty d\lambda_2 \dots \int_t^\infty d\lambda_N P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N). \quad (2.1)$$

In this paper, we will compute $F_N(t)$ using semi-classical OPs $\{\pi_k(\lambda)\}_{k \in \mathbb{N}}$ which are polynomials of the variable λ while t and a are parameters (for the sake of clarity in the notations, this dependence is omitted here):

$$\begin{cases} \langle \pi_k | \pi_{k'} \rangle = \int_t^\infty e^{-\lambda} \lambda^a \pi_k(\lambda) \pi_{k'}(\lambda) d\lambda = h_k \delta_{k,k'}, \\ \pi_k(\lambda) = \lambda^k + \zeta_k \lambda^{k-1} + \dots \end{cases} \quad (2.2)$$

The cumulative distribution $F_N(t)$ can be expressed, using standard manipulations, in terms of the norms h_k 's as

$$F_N(t) = \frac{N!}{Z_N} \prod_{k=0}^{N-1} h_k. \quad (2.3)$$

As we will see, the norms h_k 's can be computed from the three term recurrence relation satisfied by the OPs

$$\lambda \pi_k = \pi_{k+1} + S_k \pi_k + R_k \pi_{k-1}, \quad (2.4)$$

from which we deduce the following important relations:

$$R_k = \frac{h_k}{h_{k-1}}, \quad (2.5)$$

$$S_k = -t \partial_t \log h_k + 2k + a + 1, \quad (2.6)$$

$$\zeta_k = - \sum_{i=0}^{k-1} S_i. \quad (2.7)$$

Note that the starting point of our analysis is very similar to the one of Basor and Chen in Ref. [39] but the analysis of the recursion relations is different. In particular, we do not make use of ladder operators techniques, which are heavily used in Ref. [39]. In addition, we provide an asymptotic analysis of this OP system (2.2) for large N , beyond the leading order.

In section 3, we will study the variables S_k , R_k , h_k and ζ_k . In particular, we will show that S_k and R_k satisfy a coupled set of equations, named Schlesinger equations in the literature on OPs

$$\begin{cases} S_k - R_{k+1} + R_k = t \partial_t S_k, \\ 2 - S_{k+1} + S_k = t \frac{\partial_t R_{k+1}}{R_{k+1}}. \end{cases} \quad (2.8)$$

This system of equations, together with the initial condition given in Eq. (3.33) below, determines uniquely the values of R_k and S_k for all values of k , as they

can be computed by induction. It is however quite difficult to analyze the large N behavior of R_N and S_N using only this set of equations (2.8). To circumvent this difficulty, it is customary to use another set of relations, called the Laguerre-Freud equations, which we derive here using the method based on Turán determinants, as developed in Ref. [43]. They read

$$\begin{cases} R_{k+2} - R_k = S_{k+1}(2k + 4 + a + t - S_{k+1}) - S_k(2k + a + t - S_k) - 2t, \\ S_{k+1}(S_{k+1} - t) = R_{k+1}(2k + 1 + a + t - S_{k+1} - S_k) \\ \quad - R_{k+2}(2k + 5 + a + t - S_{k+2} - S_{k+1}). \end{cases} \quad (2.9)$$

Some details of this derivation have been relegated in Appendix A.

In section 4, by manipulating these two sets of equations (2.8) and (2.9), we show that $F_N(t)$, for finite N , is related to a special solution of a Painlevé V equation [see Eqs. (4.3, 4.15)], thus recovering a previous result of Tracy and Widom. Section 5 is devoted to the hard edge scaling limit:

$$N \rightarrow \infty, t \rightarrow 0, x = Nt \in \mathbb{R}^+ \text{ fixed}. \quad (2.10)$$

It is known that the limiting distribution $F_\infty(x)$ can be obtained by analyzing the large N limit of this Painlevé V equation in Eqs. (4.3, 4.15), obtained for finite N , leading to (1.8) and (1.9). As we show in section 5, the finite N corrections are then easily obtained from the Schlesinger equation (2.8), from which we obtain Eq. (1.11). Finally, section 6 is devoted to the soft edge scaling limit, when $a \sim \mathcal{O}(N)$, for N large.

3. Semi-classical Orthogonal Polynomials

To study this OP system (2.2) it is useful [23] to introduce a deformation parameter α and study the following OP system

$$\begin{cases} \langle \pi_k | \pi_{k'} \rangle = \int_t^\infty e^{-\alpha\lambda} \lambda^a \pi_k(\lambda) \pi_{k'}(\lambda) d\lambda = h_k \delta_{k,k'}, \\ \pi_k(\lambda) = \lambda^k + \dots, \end{cases} \quad (3.1)$$

such that the norms h_k 's are defined by

$$h_k = \langle \pi_k | \pi_k \rangle. \quad (3.2)$$

As we show here, some useful relations can be obtained by varying α . Eventually, we will of course set $\alpha = 1$ (2.2). The first polynomials can be computed from (3.1) to obtain

$$\pi_0(\lambda) = 1, \quad (3.3)$$

$$\pi_1(\lambda) = \lambda - \frac{1+a}{\alpha} - \frac{e^{-t\alpha} t (t\alpha)^a}{\Gamma(1+a, t\alpha)}, \quad (3.4)$$

where $\Gamma(\nu, x) = \int_x^\infty y^{\nu-1} e^{-y} dy$ is the incomplete gamma function. Obviously, the expression of the OPs π_k becomes more and more complicated as k grows. The

polynomials π_k being OPs, they satisfy a three-term recursion relation, which can be obtained as follows. As $\lambda\pi_k$ is a polynomial of degree $k+1$, we can expand it on the basis of these OPs. Because $\langle\pi_{k-2-i}|\lambda\pi_k\rangle = 0$ if $i \geq 0$, we can write the three term recurrence relation [44]:

$$\lambda\pi_k = \pi_{k+1} + S_k\pi_k + R_k\pi_{k-1}, \quad (3.5)$$

where, by definition

$$S_k h_k = \langle\pi_k|\lambda\pi_k\rangle, \quad (3.6)$$

$$R_k h_{k-1} = \langle\pi_{k-1}|\lambda\pi_k\rangle. \quad (3.7)$$

From (3.3) and (3.4), we can compute the first terms

$$\begin{cases} h_0 = \int_t^\infty e^{-\alpha\lambda} \lambda^a d\lambda = \frac{\Gamma(1+a, \alpha t)}{\alpha^{a+1}}, \\ S_0 = -\partial_\alpha \log h_0 = \frac{1+a}{\alpha} + \frac{e^{-t\alpha} t (t\alpha)^a}{\Gamma(1+a, t\alpha)}, \\ R_0 = 0, \\ \zeta_0 = 0. \end{cases} \quad (3.8)$$

3.1. Schlesinger equations

In this section, we derive a couple of recursion relations called the Schlesinger equations that couple R_k and S_k . We first write

$$R_k h_{k-1} = \langle\pi_{k-1}|\lambda\pi_k\rangle = \langle\lambda\pi_{k-1}|\pi_k\rangle. \quad (3.9)$$

Therefore, using Eq. (3.5) with the substitution $k \rightarrow k-1$, we have $\langle\lambda\pi_{k-1}|\pi_k\rangle = \langle\pi_k|\pi_k\rangle = h_k$ and finally we obtain the standard relation

$$R_k = \frac{h_k}{h_{k-1}}. \quad (3.10)$$

On the other hand, using the definition of the scalar product in (3.1), we have

$$S_k h_k = \langle\pi_k|\lambda\pi_k\rangle = \int_t^\infty e^{-\alpha\lambda} \lambda^a \lambda \pi_k^2(\lambda) d\lambda = -\partial_\alpha \langle\pi_k|\pi_k\rangle = -\partial_\alpha h_k, \quad (3.11)$$

from which we deduce the relation between S_k and h_k

$$S_k = -\partial_\alpha \log h_k. \quad (3.12)$$

By combining Eq. (3.12) and Eq. (3.10), we obtain straightforwardly

$$S_{k+1} - S_k = -\partial_\alpha \log \frac{h_{k+1}}{h_k} = -\partial_\alpha \log R_{k+1}. \quad (3.13)$$

We now study the coefficient ζ_k of the term of degree $k - 1$ in the polynomial π_k [see Eq. (2.2)]:

$$\pi_k(\lambda) = \lambda^k + \zeta_k \lambda^{k-1} + \dots \quad (3.14)$$

Taking the derivative of this equation with respect to (w.r.t.) α we obtain

$$\partial_\alpha \pi_k(\lambda) = \partial_\alpha \zeta_k \lambda^{k-1} + \dots \quad (3.15)$$

Multiplying both sides of Eq. (3.15) by λ and projecting on π_k yields

$$\partial_\alpha \zeta_k h_k = \langle \lambda \partial_\alpha \pi_k | \pi_k \rangle. \quad (3.16)$$

Besides, by looking at the term of degree k , i.e. $\propto \lambda^k$, in Eq. (3.5) we find a recursion relation between S_k and ζ_k

$$\zeta_k = \zeta_{k+1} + S_k, \quad (3.17)$$

which can be solved for ζ_k , using the initial condition (3.8), to get

$$\zeta_k = - \sum_{i=0}^{k-1} S_i. \quad (3.18)$$

Furthermore, by differentiating Eq. (3.11) w.r.t. to α we have

$$\begin{aligned} \partial_\alpha (S_k h_k) &= - \int_t^\infty e^{-\alpha \lambda} \lambda^\alpha \lambda^2 \pi_k^2(\lambda) d\lambda + 2 \int_t^\infty e^{-\alpha \lambda} \lambda^\alpha \lambda \partial_\alpha \pi_k(\lambda) \pi_k(\lambda) d\lambda \\ &= - \langle \lambda \pi_k | \lambda \pi_k \rangle + 2 \langle \lambda \partial_\alpha \pi_k | \pi_k \rangle, \end{aligned} \quad (3.19)$$

where, in the second line, we have simply used the definition of the scalar product in (2.2). Using the three-term recurrence relation (3.5) to rewrite the first term and Eq. (3.16) to rewrite the second one, we have

$$\partial_\alpha (S_k h_k) = -h_{k+1} - S_k^2 h_k - R_k^2 h_{k-1} + 2 \partial_\alpha \zeta_k h_k. \quad (3.20)$$

We can also write

$$\partial_\alpha (S_k h_k) = (\partial_\alpha S_k) h_k + S_k (\partial_\alpha h_k). \quad (3.21)$$

By replacing the left hand side of Eq. (3.20) by Eq. (3.21), dividing the resulting equation by h_k and using Eqs. (3.10) and (3.12), we obtain a last recursion relation between R_k , S_k and ζ_k

$$R_{k+1} + R_k = -\partial_\alpha S_k + 2 \partial_\alpha \zeta_k. \quad (3.22)$$

Combining Eq. (3.17) and (3.22), we have

$$R_{k+1} + R_k = \partial_\alpha \zeta_{k+1} + \partial_\alpha \zeta_k. \quad (3.23)$$

And finally, with the initial condition $R_0 = \zeta_0 = 0$ in Eq. (3.8) we obtain

$$\partial_\alpha \zeta_k = R_k. \quad (3.24)$$

Therefore substituting this relation in Eq. (3.22) we finally obtain a closed system of two coupled recursion relations between S_k and R_k

$$\begin{cases} R_{k+1} - R_k = -\partial_\alpha S_k, \\ S_{k+1} - S_k = -\partial_\alpha \log R_{k+1}, \end{cases} \quad (3.25)$$

where the second equation was previously obtained in Eq. (3.13).

Our goal now is to relate α -derivatives to t -derivatives, i.e., find a relation between $\partial_\alpha h_k$ and $\partial_t h_k$. Differentiating Eq. (3.2) w.r.t. t , using that $\langle \partial_t \pi_k | \pi_k \rangle = 0$ as $\partial_t \pi_k$ is a polynomial of degree $k-1$ (2.2), we have

$$\partial_t h_k = -e^{-\alpha t} t^a \pi_k^2(t). \quad (3.26)$$

We now start from the expression of $\partial_\alpha h_k$ given in Eq. (3.11) and use integration by parts to obtain

$$\begin{aligned} -\partial_\alpha h_k &= \int_t^\infty d\lambda \pi_k^2(\lambda) e^{-\alpha \lambda} \lambda^{a+1}, \\ &= \left[-\frac{1}{\alpha} e^{-\alpha \lambda} \lambda^{a+1} \pi_k^2(\lambda) \right]_{\lambda=t}^{\lambda=\infty} \end{aligned} \quad (3.27)$$

$$\begin{aligned} &+ \frac{1}{\alpha} \int_t^\infty d\lambda e^{-\alpha \lambda} \lambda^a \left((a+1) \pi_k^2(\lambda) + 2\lambda \pi_k(\lambda) \partial_\lambda \pi_k(\lambda) \right), \\ &= \frac{1}{\alpha} (-t \partial_t h_k + (a+1) h_k + 2 \langle \pi_k | \lambda \partial_\lambda \pi_k \rangle), \end{aligned} \quad (3.28)$$

where we have used Eq. (3.26). We can easily calculate the last scalar product

$$\langle \pi_k | \lambda \partial_\lambda \pi_k \rangle = \langle \pi_k | (k \lambda^k + \dots) \rangle = k h_k, \quad (3.29)$$

to obtain the desired relation between $\partial_\alpha h_k$ and $\partial_t h_k$

$$\alpha \partial_\alpha h_k = t \partial_t h_k - (2k + a + 1) h_k. \quad (3.30)$$

With this relation (3.30), it is then straightforward to relate $\partial_\alpha R_k$ and $\partial_\alpha S_k$ to $\partial_t R_k$ and $\partial_t S_k$, using (3.10) and (3.12). Note that from now on, we set $\alpha = 1$

$$S_k = -t \partial_t \log h_k + (2k + a + 1). \quad (3.31)$$

We finally obtain, using Eq. (3.25), the so-called Schlesinger equations [45]

$$\begin{cases} S_k - R_{k+1} + R_k = t \partial_t S_k, \\ 2 - S_{k+1} + S_k = t \partial_t \log R_{k+1}. \end{cases} \quad (3.32)$$

Note that, for $\alpha = 1$, the initial condition (3.8) reads

$$\begin{cases} h_0 = \int_t^\infty e^{-\lambda} \lambda^a d\lambda = \Gamma(1+a, t), \\ S_0 = -t\partial_t \log h_0 + a + 1 = \frac{e^{-t} t^{a+1}}{\Gamma(1+a, t)} + a + 1, \\ R_0 = 0, \\ \zeta_0 = 0. \end{cases} \quad (3.33)$$

Using the Schlesinger equations (3.32) and this initial condition (3.33), we can compute step by step all the terms for arbitrary k .

We end up this section by providing a useful relation between ζ_k [see Eq. (3.14)] and the cumulative distribution of the smallest eigenvalue $F_N(t) = \text{Prob}(\min_{1 \leq i \leq N} \lambda_i \geq t)$. One has

$$F_N(t) = \int_t^\infty d\lambda_1 \int_t^\infty d\lambda_2 \dots \int_t^\infty d\lambda_N P_{\text{joint}}(\lambda_1, \lambda_2, \dots, \lambda_N) = \frac{N!}{Z_N} \prod_{k=0}^{N-1} h_k, \quad (3.34)$$

where the last equality is obtained by using the classical tricks of replacing the Vandermonde determinant by the determinant built from the OPs π_k 's (2.2) and then use the Cauchy-Binet formula [18,19]. Using Eq. (3.18) with (3.31) as well as (3.30), we can write

$$\zeta_N = -N(N+a) + t\partial_t \log \left(\prod_{k=0}^{N-1} h_k \right). \quad (3.35)$$

And therefore, from the expression of $F_N(t)$ given in Eq. (3.34), we have

$$\zeta_N = -N(N+a) + t\partial_t \log (F_N(t)). \quad (3.36)$$

This expression thus provides a link between ζ_N , which is the first non trivial coefficient of the OP's [see Eq. (3.14)] and the cumulative distribution of the smallest eigenvalue.

3.2. Laguerre-Freud equations

In this section, we derive another set of recursion relations between the coefficients R_k 's and S_k 's, the so-called Laguerre-Freud equations, following the procedure used by Belmechi and Ronveaux in [43]. To derive these equations, we start by searching two functions Ψ and Φ , which are polynomials in λ , satisfying for any polynomial p

$$\langle \Psi | p \rangle = \langle \Phi | p' \rangle, \quad (3.37)$$

where the polynomials Φ and Ψ may depend explicitly on the parameters t and a and $p'(\lambda) \equiv \partial_\lambda p(\lambda)$. Denoting by $w(\lambda) = e^{-\lambda} \lambda^a$ the weight associate to the

scalar product in Eq. (2.2) and using an integration by parts, this relation (3.37) is satisfied provided that

$$\Psi w + (\Phi w)' = 0 \quad \text{and} \quad \Phi(\lambda = t) = 0. \quad (3.38)$$

We find that the simplest non trivial solution to this equation is given by

$$\begin{cases} \Psi(\lambda) = \lambda^2 - (2 + a + t)\lambda + t(1 + a), \\ \Phi(\lambda) = \lambda^2 - t\lambda. \end{cases} \quad (3.39)$$

Then, using this property (3.37) for $p = \pi_k^2$ and for $p = \pi_k \pi_{k+1}$, we can write after expansion, using the three-term recursion relation (3.5)

$$\begin{cases} I_{2,k} - (2 + a + t)I_{1,k} + t(1 + a)I_{0,k} = 2(J_{2,k} - tJ_{1,k}), \\ K_{2,k} - (2 + a + t)K_{1,k} + t(1 + a)K_{0,k} = L_{2,k} - tL_{1,k}, \end{cases} \quad (3.40)$$

where we have introduced the four family of integrals

$$\begin{cases} I_{m,k} = \langle \lambda^m | \pi_k^2 \rangle, \\ J_{m,k} = \langle \lambda^m | \pi_k \pi_k' \rangle, \\ K_{m,k} = \langle \lambda^m | \pi_k \pi_{k+1} \rangle, \\ L_{m,k} = \langle \lambda^m | (\pi_{k+1} \pi_k)' \rangle. \end{cases} \quad (3.41)$$

These integrals are all calculated (for $m = 0, 1$ and 2) in Appendix A. Using these expressions together with Eq. (3.10), we find the two relations

$$\begin{cases} R_{k+1} + R_k + S_k(S_k - a - t - 2k - 2) + t(2k + 1 + a) = 2 \sum_{i=0}^{k-1} S_i, \\ (S_{k+1} + S_k - 3 - a - t - 2k)R_{k+1} = 2 \sum_{i=1}^k R_i + \sum_{i=0}^k S_i^2 - t \sum_{i=0}^k S_i. \end{cases} \quad (3.42)$$

We can rewrite these equations by subtracting rank k to the rank $k + 1$ and find the two Laguerre-Freud recurrence equations (which are here of order 2):

$$\begin{cases} R_{k+2} - R_k = S_{k+1}(2k + 4 + a + t - S_{k+1}) - S_k(2k + a + t - S_k) - 2t, \\ S_{k+1}(S_{k+1} - t) = R_{k+1}(2k + 1 + a + t - S_{k+1} - S_k) \\ \quad - R_{k+2}(2k + 5 + a + t - S_{k+2} - S_{k+1}). \end{cases} \quad (3.43)$$

As we show below these two sets of equations (3.32) and (3.43) allow us to (i) derive the connection to the Painlevé equation and (ii) perform the asymptotic analysis of these coefficients for large N .

4. Painlevé V equation for finite N

In this section, we proceed to the derivation of the Painlevé V equation, following the method of Ref. [39]. First, it is useful to introduce the quantities [46]

$$\begin{cases} \theta_k = 2k + 1 + a - S_k, \\ \omega_k = -R_k - \zeta_k. \end{cases} \quad (4.1)$$

Manipulating the Laguerre-Freud equations (3.43) we can prove (see Appendix B)

$$\omega_k^2 - \theta_k(\theta_k - a - 2k + t)\omega_k - \theta_k(kt(k + a) + (\theta_k + t)\zeta_k) = 0, \quad (4.2)$$

which is a simple algebraic relation (and not a recursion relation) between the different variables ω_k , θ_k and ζ_k . Using the previous relations with the index $k = N$, we find from Eq. (3.36),

$$H_N = t\partial_t \log(F_N(t)) = N(N + a) + \zeta_N. \quad (4.3)$$

Summing up the first Schlesinger equation (3.32) from $k = 0$ to $k = N - 1$, we find

$$\sum_{k=0}^{N-1} S_k + \sum_{k=0}^{N-1} (R_k - R_{k-1}) = t\partial_t \sum_{k=0}^{N-1} S_k \quad (4.4)$$

which can be simplified using (3.18), (3.8) and (4.1) to yield

$$-t\partial_t \zeta_N = -R_N - \zeta_N = \omega_N. \quad (4.5)$$

On the other hand, using (4.3) and (4.5), we have

$$t\partial_t H_N = -\omega_N = \zeta_N + R_N = H_N - N(N + a) + R_N. \quad (4.6)$$

Taking a derivative of this equation w.r.t. t , we find

$$\partial_t R_N = t\partial_t^2 H_N. \quad (4.7)$$

From the second Schlesinger equation (3.32), we find

$$t\partial_t R_N = R_N(2 - S_N - S_{N-1}) = R_N(\theta_N - \theta_{N-1}), \quad (4.8)$$

where we have used (4.1). Using the relation derived in Appendix B in Eq. (B.9), we can rewrite this recursion equation as a simple algebraic relation

$$t\partial_t R_N = R_N\theta_N - \frac{\omega_N^2}{\theta_N}. \quad (4.9)$$

Therefore, combining (4.7) and (4.9), we find

$$t^2\partial_t^2 H_N = R_N\theta_N - \frac{\omega_N^2}{\theta_N}. \quad (4.10)$$

On the other hand, from our Eq. (4.2) in which we inject the definition of ω_N (4.1) to eliminate ζ_N

$$N(N + a)t - (2N + a)\omega_N - tR_N = R_N\theta_N + \frac{\omega_N^2}{\theta_N}. \quad (4.11)$$

Summing and subtracting the last two equations (4.10) and (4.11), we obtain

$$2\frac{\omega_N^2}{\theta_N} = N(N+a)t - (2N+a)\omega_N - tR_N - t^2\partial_t^2 H_N, \quad (4.12)$$

$$2R_N\theta_N = N(N+a)t - (2N+a)\omega_N - tR_N + t^2\partial_t^2 H_N. \quad (4.13)$$

Multiplying these two equations (4.12) and (4.13) together, we find

$$4R_N\omega_N^2 = (N(N+a)t - (2N+a)\omega_N - tR_N)^2 - (t^2\partial_t^2 H_N)^2. \quad (4.14)$$

Finally, by using (4.6), we eliminate ω_N and R_N (by expressing them in terms of H_N and $\partial_t H_N$) and find the equation

$$(t\partial_t^2 H_N)^2 = 4(\partial_t H_N)^2 (H_N - N(N+a) - t\partial_t H_N) + ((2N+a-t)\partial_t H_N + H_N)^2, \quad (4.15)$$

which is a Painlevé V equation in the Jimbo-Miwa-Okamoto σ form [47,48]. Note that this equation coincides exactly with the equation first found by Tracy and Widom in [49].

5. Large N asymptotic limit at the hard edge: Painlevé III and first correction

In this section, we study the behavior of the quantities h_N , R_N , S_N and ζ_N in the hard edge limit (2.10), which, in the language of OPs, corresponds to a double scaling limit. Of course, as we are eventually interested in the study of the cumulative distribution of the smallest eigenvalue $F_N(t)$, we could perform this asymptotic analysis directly on the Painlevé V equation (4.15), as done by Tracy and Widom in their original study of GUE [1]. But it turns out to be much more convenient, especially to extract the $1/N$ corrections, to perform this asymptotic analysis on the Schlesinger (3.32) and Laguerre-Freud (3.43) equations.

To understand the structure of the coefficients h_N , R_N , S_N and ζ_N in this double scaling limit (2.10), it is useful to study their behavior for small t (keeping N fixed). For $t = 0$, the OPs π_k 's in Eq. (2.2) can be expressed in terms of the generalized Laguerre polynomials

$$L_k^{(a)}(x) = \frac{\Gamma(a+k+1)}{k!} \sum_{i=0}^k \binom{k}{i} \frac{(-x)^i}{\Gamma(a+i+1)}. \quad (5.1)$$

Hence, thanks to standard properties of Laguerre polynomials, we easily obtain

these coefficients, for $t = 0$, using (3.2), (3.10) and (3.31) as:

$$\left\{ \begin{array}{l} \pi_k(\lambda)|_{t=0} = L_k^{(a)}(\lambda)k!(-1)^k, \\ h_k|_{t=0} = \Gamma(k+a+1)k!, \\ R_k|_{t=0} = k(k+a), \\ S_k|_{t=0} = 2k+a+1, \\ \zeta_k|_{t=0} = -k(k+a). \end{array} \right. \quad (5.2)$$

We now take the index $k = N$. When t is closed to 0, we can use Eq. (3.26) and find

$$\partial_t h_N = -t^a e^{-t} \pi_N(t)^2 = -t^a L_N^{(a)}(0)^2 N!^2 + o(t^a) = -t^a \frac{[\Gamma(N+a+1)]^2}{[\Gamma(a+1)]^2} + o(t^a). \quad (5.3)$$

Using (5.2), we can integrate and find the first correction

$$h_N = \Gamma(N+a+1)N! - \frac{t^{a+1}[\Gamma(N+a+1)]^2}{\Gamma(a+2)\Gamma(a+1)} + o(t^{a+1}). \quad (5.4)$$

We are interested in the scaling behavior when N goes to infinity, t goes to 0 with $x = Nt$ finite (2.10). In this double scaling regime, the expansion above reads

$$h_N = \Gamma(N+a+1)N! \left[1 - \frac{1}{N} \left(\frac{x^{a+1}}{\Gamma(a+2)\Gamma(a+1)} + o(x^{a+1}) \right) + o\left(\frac{1}{N}\right) \right] \quad (5.5)$$

This suggests the ansatz

$$h_N = \Gamma(N+a+1)N! \left[1 + \frac{1}{N} f(x = Nt) + o\left(\frac{1}{N}\right) \right], \quad (5.6)$$

where the function f has thus the small x expansion, read from Eq. (5.5)

$$f(x) = -\frac{x^{a+1}}{\Gamma(a+2)\Gamma(a+1)} + o(x^{a+1}). \quad (5.7)$$

We can introduce the ansatz (5.6) in the relations (3.10) and (3.31) which give R_N and S_N in terms of the function f . We obtain

$$\left\{ \begin{array}{l} R_N = N(N+a) + x f'(x) - f(x) + o(1), \\ S_N = 2N + a + 1 - \frac{1}{N} x f'(x) + o\left(\frac{1}{N}\right). \end{array} \right. \quad (5.8)$$

The first non trivial terms in Eq. (5.8) $x f'(x) - f(x)$ for R_N and $-(1/N)x f'(x)$ for S_N are necessary to compute the leading order of the cumulative distribution, $F_\infty(x)$. To compute the first $1/N$ correction to the limiting distribution, we need

to expand R_N and S_N in Eq. (5.8) to the next order in $1/N$. One actually expects the following expansion

$$\begin{cases} R_N = N(N+a) + \sum_{i=0}^j r_i(x) N^{-i} + o(N^{-j}), \\ S_N = 2N + a + 1 + \sum_{i=1}^j s_i(x) N^{-i} + o(N^{-j}), \end{cases} \quad (5.9)$$

where the first term in this expansion are given in (5.8), i.e., $r_0(x) = xf'(x) - f(x)$ and $s_1(x) = -xf'(x)$. Besides, using the exact relation $\partial_t h_N = -t^a e^{-t} \pi_N(t)^2$ [see Eq. (5.3)], one can show that $r_i(x) = o(x)$ as well as $s_i(x) = o(x)$ when $x \rightarrow 0$. One can check, in principle, the validity of this expansion (5.11) order by order in powers of $1/N$ by injecting it in the Schlesinger equations (3.32). Proving it rigorously to arbitrary order j is however a hard task. However, here, we only need this asymptotic expansion up to order $\mathcal{O}(1/N^2)$, which can be obtained explicitly as follows. To compute the second correction, we truncate the expansion (5.11) up to the N^{-2} order and write

$$\begin{cases} R_N = N(N+a) + xf'(x) - f(x) + \frac{r_1(x)}{N} + \frac{r_2(x)}{N^2} + o(N^{-2}), \\ S_N = 2N + a + 1 - \frac{xf'(x)}{N} + \frac{s_2(x)}{N^2} + o(N^{-2}). \end{cases} \quad (5.10)$$

In the hard edge limit $N \rightarrow \infty$, $t \rightarrow 0$, keeping $x = Nt$ fixed, we can obtain the expansion of R_{N+1} and S_{N+1} at the same order by replacing N by $N+1$ and using $(N+1)t = x(1+1/N)$ in (5.12)

$$\begin{cases} R_{N+1} = (N+1)(N+1+a) + xf'(x) - f(x) + \frac{1}{N}(r_1(x) + x^2 f''(x)) \\ \quad + \frac{1}{N^2}(r_2(x) + xr'_1(x) - r_1(x) + \frac{1}{2}x^2 f''(x) + \frac{1}{2}x^3 f'''(x)) + o(N^{-2}), \\ S_{N+1} = 2N + a + 3 - \frac{1}{N}xf'(x) + \frac{1}{N^2}(s_2(x) - xf'(x) - x^2 f''(x)) + o(N^{-2}). \end{cases} \quad (5.11)$$

By injecting these two expansions into the Schlesinger equation (3.32), we find (using $t\partial_t = x\partial_x$) at the first non trivial order ($\mathcal{O}(N^{-2})$ for the first Schlesinger equation and at $\mathcal{O}(N^{-3})$ for the second)

$$\begin{cases} s_2(x) - xs'_2(x) + r_1(x) - xr'_1(x) = \frac{1}{2}(x^2 f''(x) + x^3 f'''(x)), \\ 2s_2(x) - xs'_2(x) - xr'_1(x) = -ax^2 f''(x) + \frac{1}{2}x^3 f'''(x), \end{cases} \quad (5.12)$$

which can be solved as follows. First, by subtracting the first equation of (5.12) to the second one, one obtains

$$s_2(x) = r_1(x) - \left(a + \frac{1}{2}\right) x^2 f''(x). \quad (5.13)$$

By injecting this relation (5.13) in the first equation of (5.12), one finds that r_1 satisfies the following equation

$$r_1(x) - xr_1'(x) = -\frac{a}{2} (x^2 f''(x) + x^3 f'''(x)) , \quad (5.14)$$

which can be solved, using that $r_1(x) = o(x)$ as $x \rightarrow 0$, yielding

$$r_1(x) = \frac{a}{2} x^2 f''(x) . \quad (5.15)$$

Consequently, from Eq. (5.13) one has

$$s_2(x) = -\frac{a+1}{2} x^2 f''(x) . \quad (5.16)$$

We finally find the two first terms of the expansion of R_N and S_N , for large N , as

$$\begin{cases} R_N = N(N+a) + xf'(x) - f(x) + \frac{a}{2N} x^2 f''(x) + o\left(\frac{1}{N}\right), \\ S_N = 2N + a + 1 - \frac{1}{N} xf'(x) - \frac{a+1}{2N^2} x^2 f''(x) + o\left(\frac{1}{N^2}\right). \end{cases} \quad (5.17)$$

From the expansion of S_N in Eq. (5.17), we compute ζ_N in the double scaling limit (2.10), using Eq. (3.17), up to the second non-trivial order for large N

$$\zeta_N = -(N+a)N + f(x) + \frac{a}{2N} xf'(x) + o\left(\frac{1}{N}\right) . \quad (5.18)$$

Finally, from the expression of ζ_N we obtain the large N expansion of $F_N(t)$ in the hard edge limit (2.10), using Eq. (3.36), which is given by

$$x\partial_x \log\left(F_N\left(\frac{x}{N}\right)\right) = f(x) + \frac{a}{2N} xf'(x) + o\left(\frac{1}{N}\right) . \quad (5.19)$$

Using the initial condition $F_N(0) = 1$ and (5.7) we can rewrite this equation as

$$F_N\left(\frac{x}{N}\right) = \exp\left(\int_0^x \frac{f(u)}{u} du + \frac{a}{2N} f(x) + o\left(\frac{1}{N}\right)\right) . \quad (5.20)$$

Expanding Eq. (5.20) up to first order in $1/N$, one obtains finally

$$F_N\left(\frac{x}{N}\right) = F_\infty(x) + \frac{a}{2N} xF'_\infty(x) + o\left(\frac{1}{N}\right) , \quad (5.21)$$

where $F_\infty(x)$ is given by

$$\lim_{N \rightarrow \infty} F_N\left(\frac{x}{N}\right) = F_\infty(x) , \quad F_\infty(x) = \exp\left(\int_0^x \frac{f(u)}{u} du\right) . \quad (5.22)$$

Hence we easily obtain the functional form of the $1/N$ correction as given in Eq. (5.21), fully consistent with the conjecture in (1.11) made in [38]. However, at this stage, we still need to find the equation satisfied by the function f , which enters in the definition of $F_\infty(x)$ in Eq. (5.22). To obtain this equation, we expand

θ_N and ω_N in Eq. (4.1), which is easily done from the expansions of R_N , S_N and ζ_N in (5.8) and (5.18) to yield

$$\begin{cases} \theta_N = \frac{1}{N}xf'(x) + \frac{a+1}{2N^2}x^2f''(x) + o\left(\frac{1}{N^2}\right), \\ \omega_N = -xf'(x) - \frac{a}{2N}(x^2f''(x) + xf'(x)) + o\left(\frac{1}{N}\right). \end{cases} \quad (5.23)$$

Finally, by injecting these expansions (5.18) and (5.23) in (4.2) with $t = x/N$, we obtain, by canceling the first term, of order $\mathcal{O}(N^{-2})$, in Eq. (4.2) that f satisfies a Painlevé III equation

$$(xf'')^2 + 4f'(1+f')(xf' - f) = (af')^2, \quad (5.24)$$

with the small argument behavior in Eq. (5.7). This result coincides with the one obtained, by a quite different method, by Tracy and Widom [27] (note the correspondence $\sigma(s) = -f(s/4)$, where $\sigma(s)$ is the notation used in [27]). Note that for integer values of a , $f(x)$ can be written explicitly in terms Bessel functions [46] [see Eq. (C.1)]. These results in Eqs. (5.21), (5.22) and (5.24) yield the results announced in the introduction in Eqs. (1.8), (1.9) and (1.11). Finally, in Appendix C we present a comparison between numerical simulations of Wishart matrices of size $N = 50$ and the asymptotic formula in Eq. (5.21) describing the first $1/N$ correction.

6. Large N asymptotic limit at the soft edge: Painlevé II and first correction

We now turn to the analysis of the PDF of the smallest eigenvalue λ_{\min} in the case where $a \sim \mathcal{O}(N)$, and we set $a = \alpha N$. In this case, the density of eigenvalues has a single support on $[Nx_-, Nx_+]$ [see Eq. (1.4)], with $x_{\pm} = (\sqrt{1+\alpha} \pm 1)^2$ with a soft edge at both extremities (see Fig. 1). Therefore, one expects that λ_{\min} will be close to Nx_- , while its fluctuations, of order $\mathcal{O}(N^{1/3})$, are governed by the Tracy-Widom distribution for $\beta = 2$. In the soft edge limit, the large N analysis of $F_N(t)$ in Eq. (2.1) is more conveniently done directly on the Painlevé V equation (4.15) [50]. Following this route, one can indeed show [50] that, for large N

$$\lambda_{\min} = Nx_- - \frac{N^{1/3}}{m}\chi + o(N^{1/3}), \quad (6.1)$$

where m is given by [50]

$$m = \frac{(1+\alpha)^{1/6}}{(\sqrt{1+\alpha}-1)^{4/3}}, \quad (6.2)$$

and where χ is distributed according to the Tracy-Widom distribution \mathcal{F}_2 , i.e. $\Pr[\chi \leq s] = \mathcal{F}_2(s)$ where $\mathcal{F}_2(s)$ is given by [1]

$$\mathcal{F}_2(s) = \exp\left(-\int_s^\infty (x-s)q^2(x)dx\right). \quad (6.3)$$

Here $q(x)$ is the so-called Hastings-McLeod solution of the Painlevé II equation

$$q''(x) = xq(x) + 2q^3(x), \text{ with } q(x) \sim \text{Ai}(x), \text{ for } x \rightarrow \infty, \quad (6.4)$$

where $\text{Ai}(x)$ is the Airy function. The result in Eq. (6.1) can be equivalently written as

$$F_N(t) = \tilde{f}_0 \left(m \frac{Nx_- - t}{N^{1/3}} \right) + o(1), \quad (6.5)$$

which implies

$$H_N(t) = t\partial_t \log F_N(t) = -(mx_-)\tilde{h}_0(x)N^{2/3} + o(N^{2/3}), \quad x = m \frac{Nx_- - t}{N^{1/3}}. \quad (6.6)$$

with $\tilde{h}_0(x) = \tilde{f}'_0(x)/f_0(x)$ and where $\tilde{f}_0 = \mathcal{F}_2$, and where the tilde refers to the soft edge scaling limit. By injecting this form (6.6) into the Painlevé V equation satisfied by $H_N(t)$ (4.15) one finds that \tilde{h}_0 satisfies the following equation

$$\left(\tilde{h}_0''(x) \right)^2 + 4\tilde{h}_0'(x) \left[\left(\tilde{h}_0'(x) \right)^2 - x\tilde{h}_0'(x) + \tilde{h}_0(x) \right] = 0. \quad (6.7)$$

Using the Painlevé II equation (6.4), one can indeed check that $\tilde{h}_0(x) = \int_x^\infty q^2(u)du$ is solution of this equation (6.7). Note that to check this, it is useful to use the identity

$$\tilde{h}_0(x) = \int_x^\infty q^2(u)du = (q'(x))^2 - (q(x))^4 - x(q(x))^2. \quad (6.8)$$

What is the first correction to the limiting form in Eq. (6.5) ? Unfortunately, in the soft edge limit, it turns out the Schlesinger equations (3.32) do not allow to determine easily this first correction – while they were very helpful in the hard edge scaling limit. An alternative way to compute this first correction is to analyze directly the Painlevé V equation in (4.15). By inspection of this equation (4.15), we conjecture that the first correction to Eq. (6.5) is of the form

$$H_N(t) = -(mx_-) \left(\tilde{h}_0(x)N^{2/3} + \tilde{h}_1(x)N^{1/3} \right) + o(N^{1/3}). \quad (6.9)$$

By inserting this expansion (6.9) in Eq. (4.15) we obtain that \tilde{h}_1 satisfies the following linear differential equation

$$2\tilde{h}_1\tilde{h}_0' + 2(\tilde{h}_0 + h_0'(3\tilde{h}_0' - 2x))\tilde{h}_1' + \tilde{h}_0''\tilde{h}_1'' = 0, \quad (6.10)$$

where $\tilde{h}_0(x)$ is given in Eq. (6.8). Of course, we know that $\tilde{h}_1(x) \rightarrow 0$ when $x \rightarrow +\infty$ (6.9). In addition, from Eq. (6.10) one can show, using Eq. (6.8) and the large argument behavior in (6.4) $q(x) \sim \text{Ai}(x) \sim x^{-1/4}e^{-(2/3)x^{3/2}}/(2\sqrt{\pi})$ that $\tilde{h}_1(x) \sim Ax^{-1/2}e^{-(4/3)x^{3/2}}$. But, of course, the equation (6.10) being linear, the amplitude A can not be determined from this analysis. One way to determine it would be to analyze the OP system (2.2) in the limit when t is far from the left edge, i.e., for $(Nx_- - t) \gg N^{1/3}$ (corresponding to the left large deviation tail of λ_{\min} [51]), and then match this result with the typical regime, for $|Nx_- - t| \sim \mathcal{O}(N^{1/3})$. This

program was carried out in detail in a similar albeit different context, involving discrete OPs in Ref. [36]. Since we are interested in the first correction, this actually requires a very precise (and tedious) analysis of this regime $(Nx_- - t) \gg N^{1/3}$ which goes beyond the scope of the present paper. Hence our result for the first correction $\tilde{h}_1(x)$ in Eq. (6.10) does determine this function only up to a constant. We have not found any simple solution to this equation (6.10), which could indicate that the corrections to scaling in this case are actually more complicated, as found recently in the case of real Wishart matrices [35].

7. Conclusion

To conclude, we have provided a direct computation of the cumulative distribution $F_N(t)$ of the smallest eigenvalue of complex Wishart random matrices (1.1), for arbitrary parameter $a \geq 0$. This was done by studying a set of semi-classical orthogonal polynomials as defined in Eq. (2.2) for which we derived (i) the Schlesinger (2.8) and (ii) the Laguerre-Freud (2.9) equations. By combining these equations, we showed that $F_N(t)$ can be expressed in terms of a solution of a Painlevé V equation (4.15), thus recovering the result of Tracy and Widom [49] using a quite different method. In the large N limit, $F_N(t)$, properly shifted and scaled, converges to a limiting distribution $F_\infty(x)$ which can be expressed in terms of a solution of a Painlevé III equation (5.24) in the hard edge limit (corresponding to $a = \mathcal{O}(1)$) and of a Painlevé II equation (6.7) in the soft edge limit (corresponding to $a = \mathcal{O}(N)$). Furthermore, we have computed explicitly the first correction to the limiting distribution when $N \rightarrow \infty$. In the hard edge case (1.11), we confirmed a conjecture by Edelman, Guionnet et Pécché in Ref. [38]. In this case, the first correction can be simply understood as a correction to the scale of the fluctuations of λ_{\min} , see Eq. (1.12). On the other hand, in the soft edge limit, we found that this correction is a solution of a second order linear differential equation with varying coefficients (6.10). Solving this equation remains a challenging open problem, which could suggest that the first correction does not correspond to a simple shift or rescaling of the scaling variable, as found for the hard edge (1.12). This could be reminiscent of the result found for real Wishart matrices [35] and certainly deserves further investigations.

Appendix A. Some useful integrals involving the orthogonal polynomials

To compute the terms which enter the Laguerre-Freud equations, we need to compute several integrals containing the OPs $\pi_k(x)$ (2.2). To perform these computations, we will follow the method developed by Belmehdi and Ronveaux [43], using

Turán determinants. We introduce four types of integrals

$$\begin{cases} I_{m,k} = \langle \lambda^m | \pi_k^2 \rangle, \\ J_{m,k} = \langle \lambda^m | \pi_k \pi'_k \rangle, \\ K_{m,k} = \langle \lambda^m | \pi_k \pi_{k+1} \rangle, \\ L_{m,k} = \langle \lambda^m | (\pi_{k+1} \pi_k)' \rangle, \end{cases} \quad (\text{A.1})$$

for $m = 0, 1$ and 2 , while k is an arbitrary integer. Using the recurrence relation (3.5), we can calculate the integrals $I_{m,k}$'s and the $K_{m,k}$'s:

$$\begin{cases} I_{0,k} = \langle 1 | \pi_k^2 \rangle = h_k, \\ I_{1,k} = \langle \lambda | \pi_k^2 \rangle = S_k h_k, \\ I_{2,k} = \langle \lambda^2 | \pi_k^2 \rangle = (R_{k+1} + S_k^2 + R_k) h_k, \\ K_{0,k} = \langle 1 | \pi_k \pi_{k+1} \rangle = 0, \\ K_{1,k} = \langle \lambda | \pi_k \pi_{k+1} \rangle = h_{k+1}, \\ K_{2,k} = \langle \lambda^2 | \pi_k \pi_{k+1} \rangle = (S_{k+1} + S_k) h_{k+1}. \end{cases} \quad (\text{A.2})$$

The terms $J_{0,k}$, $J_{1,k}$ and $L_{0,k}$ can easily be calculated using that π_k 's are monic OPs:

$$\begin{cases} J_{0,k} = \langle 1 | \pi_k \pi'_k \rangle = 0, \\ J_{1,k} = \langle \lambda | \pi_k \pi'_k \rangle = k h_k, \\ L_{0,k} = \langle 1 | (\pi_{k+1} \pi_k)' \rangle = (k+1) h_k. \end{cases} \quad (\text{A.3})$$

The three last quantities, $J_{2,k}$, $L_{1,k}$ and $L_{2,k}$, require more work. For this purpose, we introduce a new object: the Turán determinant \mathcal{T} defined as:

$$\mathcal{T}_{k+1} = \pi_{k+2} \pi_k - \pi_{k+1}^2. \quad (\text{A.4})$$

Using the three term recurrence (3.5), we find a recurrence for the Turán determinant

$$\mathcal{T}_{k+1} = (\lambda \pi_{k+1} - S_{k+1} \pi_{k+1} - R_{k+1} \pi_k) \pi_k - \pi_{k+1} (\lambda \pi_k - S_k \pi_k - R_k \pi_{k-1}), \quad (\text{A.5})$$

$$= R_k \pi_{k+1} \pi_{k-1} + (S_k - S_{k+1}) \pi_{k+1} \pi_k - R_{k+1} \pi_k^2, \quad (\text{A.6})$$

$$= R_k \mathcal{T}_k + F_k, \quad (\text{A.7})$$

where we have introduced the auxiliary quantity $F_k = (S_k - S_{k+1}) \pi_{k+1} \pi_k + (R_k - R_{k+1}) \pi_k^2$.

Finally, we can write, using $R_k = h_k/h_{k-1}$ and the initial condition $\mathcal{T}_1 = F_0$

$$\mathcal{T}_{k+1} = h_k \sum_{i=0}^k \frac{F_i}{h_i}, \quad (\text{A.8})$$

This equation allows us to rewrite

$$\langle \lambda^m | \mathcal{T}'_{k+1} \rangle = h_k \sum_{i=0}^k h_i^{-1} ((S_i - S_{i+1})L_{m,i} + 2(R_i - R_{i+1})J_{m,i}), \quad (\text{A.9})$$

where the prime denotes a derivative w.r.t. x . On the other hand, by using the definition of the Turán determinant we find

$$\langle \lambda^m | \mathcal{T}'_{k+1} \rangle = \langle \lambda^m | (\pi_k \pi_{k+2})' \rangle - 2J_{m,k+1}. \quad (\text{A.10})$$

We obtain finally

$$\langle \lambda^m | (\pi_k \pi_{k+2})' \rangle = 2J_{m,k+1} + h_k \sum_{i=0}^k h_i^{-1} ((S_i - S_{i+1})L_{m,i} + (R_i - R_{i+1})2J_{m,i}). \quad (\text{A.11})$$

We now have all the ingredients needed to calculate the last three integrals.

Using the three term recurrence and the orthogonality condition, we can express $L_{1,k}$ as

$$L_{1,k} = \langle \lambda | \pi'_{k+1} \pi_k \rangle = S_k \langle \pi_k | \pi'_{k+1} \rangle + R_k \langle \pi_{k-1} | \pi'_{k+1} \rangle, \quad (\text{A.12})$$

$$= S_k(k+1)h_k + R_k \langle \pi_{k-1} | \pi'_{k+1} \rangle. \quad (\text{A.13})$$

From Eq. (A.11) for $m = 0$, we can express the last scalar product in terms of integrals that we have computed before. Using Eq. (A.3), we find

$$L_{1,k} = S_k(k+1)h_k + R_k h_{k-1} \sum_{i=0}^{k-1} h_i^{-1} (S_i - S_{i+1})(i+1)h_i = h_k \sum_{i=0}^k S_i. \quad (\text{A.14})$$

Using the three term recurrence and the orthogonality condition, we can express $J_{2,k}$ as

$$\begin{aligned} J_{2,k} &= \langle \lambda^2 | \pi_k \pi'_k \rangle = S_k \langle \lambda | \pi_k \pi'_k \rangle + R_k \langle \lambda | (\pi_{k-1} \pi_k)' \rangle \\ &= S_k J_{1,k} + R_k L_{1,k-1}. \end{aligned} \quad (\text{A.15})$$

Finally, we have from Eqs. (3.10), (A.3) and (A.14),

$$J_{2,k} = h_k(kS_k + \sum_{i=0}^{k-1} S_i). \quad (\text{A.16})$$

Similarly, using again the three term recurrence and the orthogonality of the family, we can express $L_{2,k}$ as

$$\begin{aligned} L_{2,k} &= \langle \lambda^2 | (\pi_{k+1} \pi_k)' \rangle \\ &= \langle \lambda^2 | \pi_{k+1} \pi'_k \rangle + \langle \lambda | \pi_{k+1} \pi'_{k+1} \rangle + S_k \langle \lambda | \pi_k \pi'_{k+1} \rangle + R_k \langle \lambda | \pi_{k-1} \pi'_{k+1} \rangle \\ &= kh_{k+1} + J_{1,k+1} + S_k L_{1,k} + R_k \langle \lambda | (\pi_{k-1} \pi_{k+1})' \rangle. \end{aligned} \quad (\text{A.17})$$

Using Eq. (A.11) for $m = 1$, we can express the last scalar product in terms of integrals that we have computed before.

$$L_{2,k} = (2k+1)h_{k+1} + S_k h_k \sum_{i=0}^k S_i + R_k 2k h_k + h_k \sum_{i=0}^{k-1} \left[(S_i - S_{i+1}) \sum_{j=0}^i S_j + 2i(R_i - R_{i+1}) \right]. \quad (\text{A.18})$$

We can simplify the telescoping sums and find

$$L_{2,k} = (2k+1)h_{k+1} + h_k \left(2 \sum_{j=1}^k R_j + \sum_{j=0}^k S_j^2 \right). \quad (\text{A.19})$$

Equipped with the computations of these integrals, we can now derive the Laguerre-Freud equations.

Appendix B. Details related to the Laguerre-Freud equations

The derivation presented here follows the one of Ref. [46] where slightly different OPs were considered. To derive the Laguerre-Freud equations, we start from equation (3.42) :

$$\begin{cases} R_{k+1} + R_k + S_k(S_k - a - t - 2k - 2) + t(2k + 1 + a) = 2 \sum_{j=1}^{k-1} S_j, \\ (S_{k+1} + S_k - 3 - a - t - 2k)R_{k+1} = 2 \sum_{j=1}^k R_j + \sum_{j=0}^k S_j^2 - t \sum_{j=0}^k S_j. \end{cases} \quad (\text{B.1})$$

It is useful to introduce the following variables

$$\begin{cases} \theta_k = 2k + 1 + a - S_k, \\ \omega_k = -R_k - \zeta_k. \end{cases} \quad (\text{B.2})$$

In terms of these variables, we can rewrite the first equation of (B.1) as

$$\omega_{k+1} + \omega_k = (t - S_k)\theta_k, \quad (\text{B.3})$$

and the second equation of (B.1) as

$$(t - S_k)(\omega_k - \omega_{k+1}) = \theta_{k-1}R_k - \theta_{k+1}R_{k+1}. \quad (\text{B.4})$$

Similarly, Eq. (B.4) can be written as

$$\begin{aligned} (\lambda + t - S_k)(\lambda + \omega_{k+1} - \omega_k) &= \theta_{k+1}R_{k+1} - \theta_{k-1}R_k + \lambda(\lambda + t - S_k + \omega_{k+1} - \omega_k) \\ &= \theta_{k+1}R_{k+1} - \theta_{k-1}R_k + \lambda(\lambda + t + R_k - R_{k+1}) \\ &= (\theta_{k+1} - \lambda)R_{k+1} - (\theta_{k-1} - \lambda)R_k + \lambda(\lambda + t). \end{aligned} \quad (\text{B.5})$$

We also rewrite Eq. (B.3) as

$$(-\lambda^2 + \lambda(2k + a - t) + 1 + \omega_{k+1} + \omega_k) = (\lambda + t - S_k)(\theta_k - \lambda). \quad (\text{B.6})$$

Finally, by multiplying together Eqs. (B.5) and (B.6), we have, with the notation $\Omega_k(\lambda) = -\frac{\lambda^2}{2} + \frac{\lambda}{2}(2k + a - t) + \omega_k$,

$$(\Omega_{k+1}(\lambda) - \Omega_k(\lambda))(\Omega_{k+1}(\lambda) + \Omega_k(\lambda)) = (\theta_k - \lambda)((\theta_{k+1} - \lambda)R_{k+1} - (\theta_{k-1} - \lambda)R_k + \lambda(\lambda + t)). \quad (\text{B.7})$$

For $\lambda = 0$, we find

$$\omega_{k+1}^2 - \omega_k^2 = R_{k+1}\theta_{k+1}\theta_k - R_k\theta_k\theta_{k-1}. \quad (\text{B.8})$$

Using the initial condition $R_0 = \omega_0 = 0$, we can solve this equation and find

$$\omega_k^2 = R_k\theta_k\theta_{k-1}. \quad (\text{B.9})$$

For $\lambda = -t$, we find

$$\Omega_{k+1}(-t)^2 - \Omega_k(-t)^2 = (\theta_k + t)(\theta_{k+1} + t)R_{k+1} - (\theta_{k-1} + t)(\theta_k + t)R_k. \quad (\text{B.10})$$

Again, using the initial condition $R_0 = \omega_0 = 0$, we can solve this equation and find

$$\Omega_k(-t)^2 = (\theta_{k-1} + t)(\theta_k + t)R_k + \left(\frac{t}{2}(2k + a)\right)^2. \quad (\text{B.11})$$

Finally, using the explicit expression of $\Omega_k(-t)$, this equation can also be written as

$$(\omega_k - kt)(\omega_k - (k + a)t) = R_k(\theta_k + t)(\theta_{k-1} + t). \quad (\text{B.12})$$

Thanks to Eq. (B.9), we can substitute θ_{k-1} in (B.12). Using the definition of ω_k in (B.2), Eq. (B.12) yields a quadratic equation for ω_k in terms of θ_k and ζ_k

$$\omega_k^2 - \theta_k(\theta_k - a - 2k + t)\omega_k - \theta_k(kt(k + a) + (\theta_k + t)\zeta_k) = 0, \quad (\text{B.13})$$

which yields the equation given in the text in Eq. (4.2).

Appendix C. Numerical simulations

In this section, we present a numerical check of our formula for the first finite N corrections to the limiting density of λ_{\min} in the hard edge case, i.e., keeping a finite while $N \rightarrow \infty$. Our simulations have been carried out for integer values of the parameter a for which the function $f(x)$ in Eq. (5.22) can be computed explicitly in terms of Bessel I functions [28]:

$$f(x) = -x \frac{\det[I_{j-k+2}(2\sqrt{x})]_{1 \leq j, k \leq a}}{\det[I_{j-k}(2\sqrt{x})]_{1 \leq j, k \leq a}}, \quad a \in \mathbb{N}^*, \quad (\text{C.1})$$

and $f(x) = -x$ for $a = 0$. In Fig. (2) we show a plot of $\log F_N(x/N)$ obtained by sampling 3.10^7 independent Wishart matrices of size $N = 50$ (dots) for different values of $a = 0, 1, 2$ and 3. The comparison with $\log F_\infty(x)$, evaluated exactly

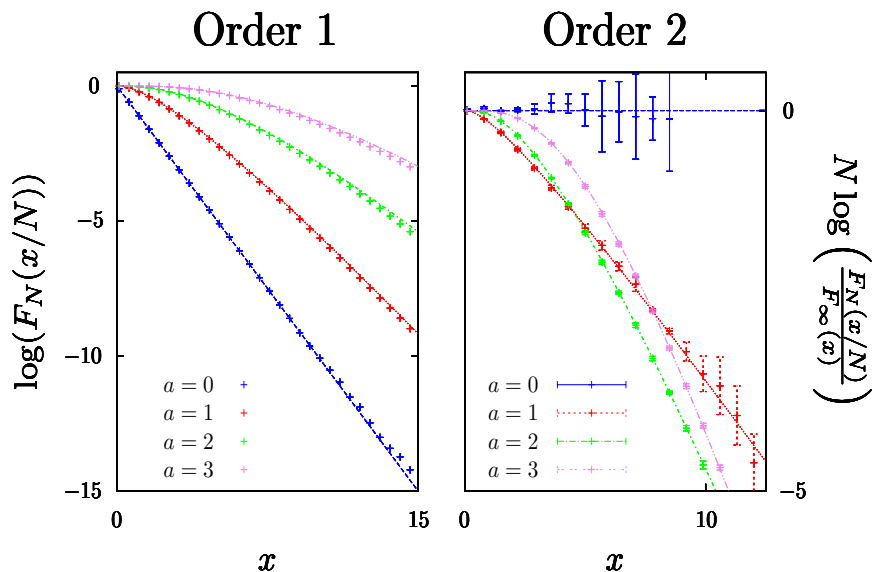


Figure 2. Comparison between our formula (5.20) using (C.1) (dashed curves) and data obtained by sampling 3.10^7 independent Wishart matrices of size $N = 50$ (dots) for different values of $a = 0, 1, 2$ and 3 . The left panels presents a comparison between $\log(F_N(x/N))$ and $\log(F_\infty(x))$ given in (5.22). The right panel illustrates the finite N correction by comparing $N \log \left(\frac{F_N(x/N)}{F_\infty(x)} \right)$ to $\frac{a}{2}f(x)$, see Eq. (C.2).

from Eqs. (5.22) and (C.1) shows a good agreement between theory and numerical simulations. The main objective of these simulations is to test the formula for the first order correction, given in Eq. (5.21) which can also be written as

$$\log \left(\frac{F_N(x/N)}{F_\infty(x)} \right) = \frac{1}{N} \frac{a}{2} f(x) + o(1/N). \quad (\text{C.2})$$

In the left panel of Fig. 2, we show a plot of $N \log [F_N(x/N)/F_\infty(x)]$ as a function of x and compare it to $\frac{a}{2}f(x)$, where $f(x)$ is given in Eq. (C.1). This comparison illustrates the accuracy of the asymptotic formula characterizing the finite N correction to $F_\infty(x)$.

Bibliography

- [1] C. A. Tracy, H. Widom, *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys., **159**, 151 (1994).
- [2] C. A. Tracy, H. Widom, *On the orthogonal and symplectic ensembles*, Commun. Math. Phys., **177**, 727 (1996).
- [3] For a review see S. N. Majumdar, Les Houches lecture notes on Complex Systems (2006), ed. by J.-P. Bouchaud, M. Mézard and J. Dalibard [arXiv: cond-

- mat/0701193]. *Random matrices, the ulam problem, directed polymers & growth models, and sequence matching*
- [4] J. Baik, P. Deift, K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Am. Math. Soc. **12**, 1119 (1999).
 - [5] M. Kardar, G. Parisi, Y.C. Zhang. *Dynamic scaling of growing interfaces*, Phys. Rev. Lett., **56**, 889 (1986).
 - [6] K. Johansson. *Shape fluctuations and random matrices*, Commun. Math. Phys., **209**(2), 437 (2000).
 - [7] M. Prähofer, H. Spohn, *Universal distributions for growth processes in $1+1$ dimensions and random matrices*, Phys. Rev. Lett. **84**, 4882 (2000).
 - [8] T. Sasamoto, H. Spohn, *One-dimensional Kardar-Parisi-Zhang equation: an exact solution and its universality*, Phys. Rev. Lett. **104**, 230602 (2010).
 - [9] P. Calabrese, P. Le Doussal, A. Rosso, *Free-energy distribution of the directed polymer at high temperature*, Europhys. Lett. **90**, 20002 (2010).
 - [10] V. Dotsenko, *Bethe ansatz derivation of the Tracy-Widom distribution for one-dimensional directed polymers*, Europhys. Lett. **90**, 20003 (2010).
 - [11] G. Amir, I. Corwin, J. Quastel, *Probability distribution of the free energy of the continuum directed random polymer in $1+1$ dimensions*, Comm. Pure and Appl. Math. **64**, 466 (2011).
 - [12] S. N. Majumdar, S. K. Nechaev, *Exact Asymptotic Results for the Bernoulli Matching Model of Sequence Alignment*, Phys. Rev. E **72**, 020901(R) (2005).
 - [13] C. Nadal, S. N. Majumdar, *Nonintersecting Brownian interfaces and Wishart random matrices*, Phys. Rev. E, **79**, 061117 (2009).
 - [14] P. J. Forrester, S. N. Majumdar, G. Schehr, *Non-intersecting Brownian walkers and Yang-Mills theory on the sphere*, Nucl. Phys. B **844**, 500 (2011).
 - [15] K. Liechty, *Nonintersecting Brownian motions on the half-line and discrete Gaussian orthogonal polynomials*, J. Stat. Phys. **147**, 582 (2012).
 - [16] G. Biroli, J.-P. Bouchaud, M. Potters, *On the top eigenvalue of heavy-tailed random matrices*, Eur. Phys. Lett. **78**, 10001 (2007).
 - [17] For a review see S. N. Majumdar, G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, J. Stat. Mech. P01012 (2014)
 - [18] M. L. Mehta. *Random Matrices*, Academic Press, 2nd Edition (1991).
 - [19] P. J. Forrester. *Log-gases and random matrices*, Princeton University, Press, Princeton, NJ (2010).
 - [20] C. A. Tracy, H. Widom. *Nonintersecting brownian excursions*, The Annals of Applied Probab., **17**(3), 953-979 (2007).
 - [21] G. Schehr, S. N. Majumdar, A. Comtet, J. Randon-Furling. *Exact distribution of the maximal height of p vicious walkers*, Phys. Rev. Lett., **101**(15), 150601 (2008).
 - [22] I. M. Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Ann. Statist. **29**, 295 (2001).
 - [23] C. Nadal, S. N. Majumdar. *A simple derivation of the Tracy-Widom distribution of the maximal eigenvalue of a Gaussian unitary random matrix*, J. Stat. Mech., P04001 (2011).
 - [24] Y. Chen, G. Pruessner, *Orthogonal polynomials with discontinuous weights*, J. Phys. A: Math. Gen. **38**(12), L191 (2005).
 - [25] P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nucl. Phys. B **402**(3), 709 (1993).
 - [26] J. J. M. Verbaarschot and I. Zahed, *Spectral density of the QCD Dirac operator near zero virtuality*, Phys. Rev. Lett. **70**, 3852 (1993).
 - [27] C. A. Tracy, H. Widom. *Level spacing distributions and the Bessel kernel*, Commun.

- Math. Phys., **161**(2), 289-309 (1994).
- [28] P. J. Forrester, T. D. Hughes. *Complex Wishart matrices and conductance in mesoscopic systems: exact results*, J. Math. Phys., **35**(12), 6736 (1994).
 - [29] G. Akemann, T. Guhr, M. Kieburg, R. Wegner, T. Wirtz, *Completing the picture for the smallest eigenvalue of real Wishart matrices*, Phys. Rev. Lett. **113**, 250201 (2014).
 - [30] G. Györgyi, N. R. Moloney, K. Ozogány, Z. Rácz, M. Droz, *Renormalization group theory for finite-size scaling in extreme statistics*, Phys. Rev. E **81**, 041135 (2010).
 - [31] G. Schehr, S. N. Majumdar, *Universal Asymptotic Statistics of Maximal Relative Height in One-dimensional Solid-on-solid Models*, Phys. Rev. E **73**, 056103 (2006).
 - [32] N. El Karoui, *A rate of convergence result for the largest eigenvalue of complex white Wishart matrices*, Ann. Probab. **34**, 2077 (2006).
 - [33] I. M. Johnstone, *Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence*, Ann. Statist. **36**, 2638 (2008).
 - [34] I. M. Johnstone and Z. Ma, *Fast approach to the Tracy-Widom law at the edge of GOE and GUE*, Ann. Appl. Probab. **5**, 1962 (2012).
 - [35] Z. Ma, *Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices*, Bernoulli **18**, 322 (2012).
 - [36] G. Schehr, S. N. Majumdar, A. Comtet, P. J. Forrester, *Reunion probability of N vicious walkers: typical and large fluctuations for large N* , J. Stat. Phys. **150**, 491 (2013).
 - [37] J. Baik, R. Jenkins, *Limiting distribution of maximal crossing and nesting of Poissonized random matchings*, Ann. Probab. **41**, 4359 (2013).
 - [38] A. Edelman, A. Guionnet, S. Péché. *Beyond Universality in Random Matrix Theory*, preprint arXiv, 1405.7590 (2014).
 - [39] E. Basor, Y. Chen, *Painlevé V and the distribution function of a discontinuous linear statistic in the Laguerre unitary ensembles*, J. Phys. A: Math. Theor. **42**, 035203 (2009).
 - [40] G. Schehr, Talk given at FOCM'14, December 16, 2014, http://lptms.u-psud.fr/gregory-schehr/files/2014/12/Talk_FOCM.pdf, (2014).
 - [41] F. Bornemann, *A note on the expansion of the smallest eigenvalue distribution of the LUE at the hard edge*, preprint arXiv:1504.00235.
 - [42] W. Hachem, A. Hardy, J. Najim, *A Survey on the Eigenvalues Local Behavior of Large Complex Correlated Wishart Matrices*, to appear (2015).
 - [43] S. Belmehdi, A. Ronveaux. *Laguerre-Freud's Equations for the Recurrence Coefficients of Semi-classical Orthogonal Polynomials*, J. Approx. Theory, **76**(3), 351-368 (1994).
 - [44] G. Szegő. *Orthogonal polynomials*, American Math. Society, Providence R.I., 4th edition (1975).
 - [45] A. P. Magnus. *Painlevé-type differential equations for the recurrence coefficients of semi-classical orthogonal polynomials*, J. Comput. Appl. Math., **57**(1), 215 (1995).
 - [46] P. J. Forrester, N. S. Witte. *The distribution of the first eigenvalue spacing at the hard edge of the Laguerre unitary ensemble*, Kyushu Journ. Math **61**, 457 (2007).
 - [47] M. Jimbo, T. Miwa, *Monodromy preserving deformation of linear ordinary differential equations with rational coefficients: II*, Physica D **2**, 407 (1981).
 - [48] K. Okamoto, *On the τ -function of the Painlevé equations*, Physica F **2**, 525 (1982).
 - [49] C. A. Tracy, H. Widom, *Fredholm determinants, differential equations and matrix models*, Commun. Math. Phys. **163**, 33 (1994).
 - [50] T. H. Baker, P. J. Forrester, P. A. Pearce. *Random matrix ensembles with an effective extensive external charge*, J. Phys. A, **31**, 6087 (1998).

- [51] E. Katzav, I.P. Castillo. *Large deviations of the smallest eigenvalue of the Wishart-Laguerre ensemble*, Phys. Rev. E, **82**(4), 040104 (2010).